

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 4 年 1 月 3 0 日
Date of Application:

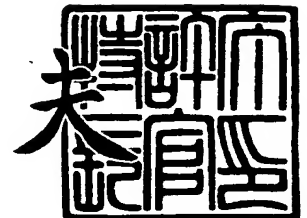
出 願 番 号 特 願 2 0 0 4 - 0 2 3 0 9 0
Application Number:
[ST. 10/C] : [J P 2 0 0 4 - 0 2 3 0 9 0]

出 願 人 株式会社日立製作所
Applicant(s):

2 0 0 4 年 3 月 8 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



【書類名】 特許願
【整理番号】 NT03P0847
【提出日】 平成16年 1月30日
【あて先】 特許庁長官 殿
【国際特許分類】 G06F 12/00
G06F 12/00520

【発明者】
【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
システム開発研究所内
【氏名】 川村 俊二

【発明者】
【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
システム開発研究所内
【氏名】 江口 賢哲

【特許出願人】
【識別番号】 000005108
【氏名又は名称】 株式会社日立製作所

【代理人】
【識別番号】 100068504
【弁理士】
【氏名又は名称】 小川 勝男
【電話番号】 03-3537-1621

【選任した代理人】
【識別番号】 100086656
【弁理士】
【氏名又は名称】 田中 恭助
【電話番号】 03-3537-1621

【選任した代理人】
【識別番号】 100094352
【弁理士】
【氏名又は名称】 佐々木 孝
【電話番号】 03-3537-1621

【手数料の表示】
【予納台帳番号】 081423
【納付金額】 21,000円

【提出物件の目録】
【物件名】 特許請求の範囲 1
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1

【書類名】 特許請求の範囲**【請求項 1】**

第一の計算機と前記第一の計算機に接続された前記第一の記憶装置システムとを備えるプライマリサイトと、第二の計算機と前記第二の計算機に接続された前記第二の記憶装置システムとを備えるセカンダリサイトとを有するデータ処理システムにおいて、

前記第一の記憶装置システムと前記第二の記憶装置システムとは、通信線により接続され、

前記第一の記憶装置システムは、データの更新履歴をジャーナルとして記憶装置に記録し、前記ジャーナルを第二の記憶装置システムに前記通信線を介して転送し、

前記第二の記憶装置システムは、転送されてきた前記ジャーナルを記憶装置に格納することを特徴とするデータ処理システム。

【請求項 2】

前記第二の記憶装置システムは、格納された前記ジャーナルに基づいて、データ復元をおこなうことを特徴とする請求項 1 記載のデータ処理システム。

【請求項 3】

前記第一の記憶装置システムの記憶装置に記録された前記ジャーナルは、複数の論理ボリュームに格納され、

ある論理ボリュームに格納中に、他の論理ボリュームに格納用の論理ボリュームを切り替えることを特徴とする請求項 1 記載のデータ処理システム。

【請求項 4】

前記切り替えのタイミングが、ユーザの指定したタイミングであることを特徴とする請求項 3 記載のデータ処理システム。

【請求項 5】

前記切り替えのタイミングが、前記転送の対象としているジャーナルが格納されている論理ボリュームが前記セカンダリサイトに転送され終わったタイミングであることを特徴とする請求項 3 記載のデータ処理システム。

【請求項 6】

前記第二の記憶装置システムの転送されたきた前記ジャーナルは、複数の論理ボリュームに格納され、

ある論理ボリュームに転送中に、他の論理ボリュームに転送先の論理ボリュームを切り替えることを特徴とする請求項 1 記載のデータ処理システム。

【請求項 7】

前記切り替えのタイミングが、ユーザの指定したタイミングであることを特徴とする請求項 6 記載のデータ処理システム。

【請求項 8】

前記第二の記憶装置システムは、格納された前記ジャーナルに基づいて、データ復元をおこない、かつ、

前記切り替えのタイミングが、復元に用いているジャーナルを格納している論理ボリュームが全て復元されたタイミングであることを特徴とする請求項 6 記載のデータ処理システム。

【請求項 9】

前記第二の記憶装置システムは、前記第一の記憶装置システムにおいて記録されたジャーナルに関する情報を取得し、

前記第二の記憶装置システムは、前記第一の記憶装置システムに対して、前記ジャーナルの送付を要求するコマンドを発行することを特徴とする請求項 1 記載のデータ処理システム。

【請求項 10】

前記第一の記憶装置システムの記憶装置に記録された前記ジャーナルは、複数の論理ボリュームに格納され、

ある論理ボリュームに格納中に、他の論理ボリュームに格納用の論理ボリュームを切り

替え、

前記切り替えのタイミングが、前記第二の記憶装置システムから、前記ジャーナルの送付を要求するコマンドを受信したタイミングであることを特徴とする請求項 9 記載のデータ処理システム。

【請求項 11】

前記第二の記憶装置システムの転送されたきた前記ジャーナルは、複数の論理ボリュームに格納され、

ある論理ボリュームに転送中に、他の論理ボリュームに転送先の論理ボリュームを切り替え、

前記切り替えのタイミングが、前記第一の記憶装置システムから、前記ジャーナルの転送が開始されたタイミングであることを特徴とする請求項 9 記載のデータ処理システム。

【請求項 12】

第一の計算機と前記第一の計算機に接続された記憶装置システムとを備えるプライマリサイトと、第二の計算機と前記第二の計算機に接続された記憶装置システムとを備えるセカンダリサイトとを有するデータ処理システムにおいて、

前記第一の計算機と前記第二の計算機とは、第一の通信線により接続され、

前記第二の記憶装置システムと前記第二の記憶装置システムとは、第二の通信線により接続され、

前記第一の記憶装置システムは、データの更新履歴をジャーナルとして記憶装置に記録し、

前記第一の計算機は、前記第一の記憶装置システムから前記ジャーナルに関する情報を取得して、前記第一の通信線を介して前記第二の記憶装置システムに送信し、

前記第一の記憶装置システムは、前記ジャーナルを第二の記憶装置システムに前記第二の通信線を介して転送し、

前記第二の記憶装置システムは、転送されてきた前記ジャーナルを記憶装置に格納することを特徴とするデータ処理システム。

【請求項 13】

前記第二の記憶装置システムは、前記第一の記憶装置システムに対して、前記ジャーナルの送付を要求するコマンドを発行することを特徴とする請求項 12 記載のデータ処理システム。

【請求項 14】

前記第二の記憶装置システムにおけるデータの復元は、前記転送されたジャーナルに基づき、前記第二の計算機で実行される復元プログラムによりおこなわれることを特徴とする請求項 12 記載のデータ処理システム。

【請求項 15】

第一の計算機と前記第一の計算機に接続された前記第一の記憶装置システムとを備えるプライマリサイトと、第二の計算機と前記第二の計算機に接続された前記第二の記憶装置システムとを備えるセカンダリサイトとを有するデータ処理システムにおいて、

前記第一の記憶装置システムと前記第二の記憶装置システムとは、通信線により接続され、

前記第一の記憶装置システムは、第一の記憶制御装置と第一の記憶装置とを有し、

前記第一の記憶制御装置は、データの更新履歴をジャーナルとして前記第一の記憶装置に記録するジャーナル取得プログラムと、前記ジャーナルを第二の記憶装置システムに前記通信線を介して転送するジャーナル転送プログラムとを実行し、

前記第二の記憶装置システムは、第二の記憶制御装置と第二の記憶装置とを有し、

前記第二の記憶装置システムは、データをジャーナルに基づいて復元するジャーナル復元プログラムと、前記ジャーナルを第一の記憶装置システムから転送された受け取るジャーナル転送プログラムとを実行し、

前記第一の記憶装置システムから前記第二の記憶装置システムに前記ジャーナル転送時に、

前記第一の記憶制御装置は、前記第一の記憶装置のある論理ボリュームに格納中に、前記第一の記憶装置の他の論理ボリュームに格納用の論理ボリュームを切り替え、

前記第二の記憶制御装置は、前記第二の記憶装置のある論理ボリュームに転送中に、前記第二の記憶装置の他の論理ボリュームに転送先の論理ボリュームを切り替えることを特徴とするデータ処理システム。

【書類名】明細書

【発明の名称】データ処理システム

【技術分野】

【0001】

本発明は、データ処理システムに係り、特に、複数のサイトにジャーナルを分散して保持する技術に用いて好適なデータ処理システムに関する。

【背景技術】

【0002】

データ処理システムにおいて災害等による記憶装置システム内のデータ損失を避けることが要請されている。そのために、遠隔地の記憶装置システムにデータを複製する技術が重要になっている。ここで、記憶装置システムは、記憶制御装置とディスク装置などの記憶装置を含むものとする。

【0003】

記憶装置システムに格納されたデータを、別の記憶装置システムに複製する技術としては、特許文献1に開示された技術がある。特許文献1には、第一システムに含まれる計算機（以下、「プライマリホスト」という）のOSのデバイスドライバがライトシステムコールを受けると、データをローカルデータデバイスに書き込むとともに更新ログをライトログデバイスに格納し、プライマリホストのプログラムが第二のシステムに含まれる計算機（以下、「セカンダリホスト」という）のプログラムに更新ログ情報を転送し、セカンダリホストのプログラムが受け取った更新ログ情報を元に第二のシステムのデータデバイスのデータを更新する技術が開示されている。

【0004】

【特許文献1】米国特許6324654号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

上記従来技術では、プライマリホストとセカンダリホストの間で双方の記憶装置システムに格納されたデータの転送がおこなわれる。このときに、ホストがデータ転送の経路として使用される。そして、各記憶装置システムに格納されたデータがホスト間の通信リンクを介して転送されるため、各ホストのCPU負荷、チャネル負荷、および、ホスト間を接続する回線のトラフィックが増加するという問題点があった。また、ログを採取することによるデバイスの負荷分散について考慮されていないため、ログ書き込み処理とログ読み出し処理がライトログデバイスへ集中するという問題点があった。

【0006】

本発明は、上記従来技術の問題点を解決するためになされたもので、その目的は、データの更新記録を採取してデータの復元をおこなうことのできるデータ処理システムにおいて、ホストとネットワークの負荷をかけることなく、しかも、データの更新復元に伴う特定の記憶装置への負荷集中を避けて、システムの処理能力が低下せずに、複数のサイトでデータの整合性を保証できるデータ処理システムを提供することにある。

【課題を解決するための手段】

【0007】

本発明のデータ処理システムは、プライマリサイトとセカンダリサイトで構成されたシステムであって、各サイトには、ホストと記憶装置システムを備えている。

【0008】

そして、プライマリサイトの記憶装置システム（以下、「プライマリ記憶装置システム」という）に格納されたデータの更新の情報をジャーナル（更新履歴）として格納する。ジャーナルは、具体的には、更新に用いられたデータのコピーとメタデータの記録である。

【0009】

プライマリ記憶装置システムは、セカンダリサイトの記憶装置システム（以下、「セカ

ンダリ記憶装置システム」という)と接続された通信線を介して、このジャーナルをセカンダリ記憶装置システムに転送する。セカンダリ記憶装置システムは、プライマリ記憶装置システムから受け取ったジャーナルを用いてセカンダリ記憶装置システムに格納されたデータを復元する(以下では、ジャーナルによりデータ復元をおこなうことを、「ジャーナル復元」ともいうことにする)。

【0010】

本発明では、単にデータではなく、ジャーナルをセカンダリサイトに転送して復元をおこなうので、障害が起こっても任意の時点のデータを速やかに復旧でき、データの整合性を保証することができる。

【0011】

また、本発明においては、プライマリ記憶装置システムは、ジャーナルを格納している論理ボリュームを複数持ち、現時点においてジャーナルの記録に用いている格納用論理ボリュームを切り替えることによつて、ジャーナルの転送に用いるジャーナル転送元の論理ボリュームとアクセスの集中を避け、負荷分散を図ることができる。

【0012】

同様にセカンダリ記憶装置システムではジャーナル転送に用いるジャーナル転送先の論理ボリュームを切り替え、ジャーナル復元に用いている論理ボリュームと別ボリュームとすることでアクセスの集中を避け、負荷分散を図ることができる。

【0013】

なお、ジャーナルの転送は、セカンダリ記憶装置システムがプライマリ記憶装置システムにジャーナル転送要求を発行することで実現する構成としても良い。

【0014】

また、プライマリホストおよびセカンダリホストが、各ホスト上で動作するプログラムに基づいて、各ホストに接続された記憶装置システムの状態を監視しておき、必要に応じて、それぞれのサイトのホストが、記憶装置システム間のデータ転送を、そのサイトの記憶装置システムに指示する構成としても良い。

【発明の効果】

【0015】

本発明によれば、データの更新記録を採取してデータの復元をおこなうことのできるデータ処理システムにおいて、ホストとネットワークの負荷をかけることなく、しかも、データの更新・復元に伴う特定の記憶装置への負荷集中を避けて、システムの処理能力の低下が少なく、複数のサイトでデータの整合性を保証できるデータ処理システムを提供することができる。

【発明を実施するための最良の形態】

【0016】

以下、本発明に係る各実施形態を、図1ないし図17を用いて説明する。

【0017】

〔実施形態1〕

以下、本発明に係る第一の実施形態を、図1ないし図13を用いて説明する。

(I) データ処理システムの構成

先ず、図1を用いて本発明の第一の実施形態に係るデータ処理システムの構成について説明する。

図1は、本発明の第一の実施形態に係るデータ処理システムのハードウェア構成を示す図である。

図2は、本発明の第一の実施形態に係るデータ処理システムの機能構成を示す図である。

【0018】

本実施形態のデータ処理システムは、各サイトが連携して処理をおこない、各サイトには、ホストと記憶装置システムを備えている。

【0019】

ここで、第一のサイトを「プライマリサイト」といい、第二のサイトを「セカンダリサイト」といい、プライマリサイトからセカンダリサイトにジャーナルを転送する例を説明する。また、プライマリサイトに属するホストを「プライマリホスト」、プライマリサイトに属する記憶装置システムを「プライマリ記憶装置システム」、プライマリ記憶装置システムの記憶装置に格納されるジャーナルを、「プライマリジャーナル」、プライマリジャーナルを格納するためのボリュームを「プライマリジャーナルボリューム」ということにする。セカンダリサイトについても同様である。

【0020】

さて、図1に示されるように、プライマリサイト100Aは、プライマリホスト110A、プライマリ記憶装置システム120Aを有していて、セカンダリサイト100Bは、セカンダリホスト110B、セカンダリ記憶装置システム120Bを有していて、それらに管理端末130が接続されている。

【0021】

各ホスト110（プライマリホスト110A、セカンダリホスト110B）は、プロセッサ111、主記憶装置112、および、入出力処理装置113を有する計算機である。具体的には、ワークステーション、パーソナルコンピュータ、メインフレーム等である。

【0022】

各記憶装置システム120は、記憶制御装置140、一つ以上の記憶装置121および、保守端末122を有する。記憶装置121は、磁気ディスク記憶装置や光ディスク記憶装置などの補助記憶装置である。また、保守端末122は備えていないこともある。

【0023】

記憶制御装置140は、ホスト入出力処理装置141、キャッシュメモリ142、ディスク入出力処理装置143、プロセッサ144、および、制御メモリ145を有する。

【0024】

ホストは、記憶装置システムとLAN（Local Area Network）やSAN（Storage Area Network）といったホスト-記憶装置システム間ネットワーク150によって互いに接続される。各ホストのプロセッサ111と主記憶装置112は、入出力処理装置113とホスト-記憶装置システム間ネットワーク150を介して記憶装置システム120のホスト入出力処理装置141に接続される。

【0025】

記憶装置システム同士は、記憶装置システム間ネットワーク160を介して接続される。記憶装置システム間ネットワーク160は、一般に公衆回線などのグローバルネットワークである場合が多く、通信サービス提供者から有料で借用することが多い。システムの安全性のために（両サイトが同時に障害にあわないように）距離をおくためにグローバルネットワークを使用することが多いが、記憶装置システム同士が同じ部屋の中、同じビルの中、近隣ビルの中にある場合などはローカルネットワークを使用してもよい。ただし、このようなネットワークの形態によって本発明が限定されるものではない。

【0026】

管理端末130も、プロセッサや主記憶装置を有する計算機である。管理端末130、プライマリホスト110A、セカンダリホスト110B、プライマリ記憶装置システム120A、および、セカンダリ記憶装置システム120Bは、LANやWAN等のネットワーク170を介して相互に接続される。

【0027】

このようなデータ処理システムの機能的な構成を示すと図2のようになる。

【0028】

各記憶制御装置140では、記憶装置システム120間のデータ転送を制御するプログラムであるジャーナル処理管理プログラム221、ジャーナル転送プログラム222、ジャーナル取得・復元プログラム223がプロセッサ144上で実行される。これらのプログラムは制御メモリ145に格納される。

【0029】

ジャーナル処理管理プログラム 221 は各記憶装置システムが実行するジャーナル処理（ジャーナル取得処理、ジャーナル転送処理、ジャーナル復元処理）の管理をおこなう。詳細は後述する。また、ジャーナル処理中は各記憶制御装置 140 のジャーナル処理管理プログラム 221 間で随時通信がおこなわれ、ジャーナル処理に必要な管理情報を交換する。

【0030】

ジャーナル取得・復元プログラム 223 は、プロセッサ 144 にジャーナルの取得とジャーナルの復元を実行させるプログラムであり、ジャーナル取得プログラムとジャーナル復元プログラムから構成される。

【0031】

また、記憶制御装置 140 は、このようなジャーナル処理管理プログラム 221、ジャーナル転送プログラム 222、ジャーナル取得・復元プログラム 223 に関する処理の他にも、各ホストからの指示に基づいて、記憶装置 121 への入出力処理を実行する。

【0032】

記憶装置 121 には、一つ以上の論理記憶領域（論理ボリューム）が作成される。論理ボリュームは、記憶装置 121 が有する物理的な記憶領域と関連付けられる。これらの論理ボリュームはユーザの指定によりデータボリューム領域 225 とジャーナルボリューム領域 227 として使用される。なお、各ホスト 110 では、ユーザが使用するアプリケーションプログラム 211 や、記憶装置システムとのインタフェース制御をおこなう記憶装置制御プログラム 212 もホスト 110 が有するプロセッサ 111 で実行される。また、ジャーナル処理管理プログラムと記憶装置制御プログラム 212 は相互に情報をやり取りする。

【0033】

なお、データ複製のコピー元となるプライマリ記憶装置システム 120A が有するデータボリュームを、以下「PVOL」といい、PVOL に格納されるデータの複製先となるセカンダリ記憶装置システム 120B が有するデータボリュームを、以下「SVOL」ということにする。

【0034】

管理端末 130 で実行される管理プログラム 231 は、本実施形態のデータ処理システムの各構成要素、具体的には各ホスト 110 や各記憶装置システム 120 の管理をおこなう。管理プログラム 231 は管理端末 130 の主記憶装置に格納される。

【0035】

なお、ここまで説明した各プログラムは、コンパクトディスクや光磁気ディスクといった可搬媒体を用いて、あるいは、ネットワーク 170 を介して、各装置が有する記憶媒体にインストールされる。

(II) データ処理システムに用いられるデータ構造

次に、図 3 ないし図 9 を用いて本実施形態のデータ処理システムに用いられるデータ構造について説明する。

図 3 は、ジャーナルグループ管理情報を示す図である。

図 4 は、データボリューム管理情報を示す図である。

図 5 は、ジャーナルメタ情報を示す図である。

図 6 は、データボリュームとジャーナルボリュームの対応を示す図である。

図 7 は、プライマリジャーナルボリューム内のジャーナルデータ領域 820 の内部構造を示す図である。

図 8 は、セカンダリジャーナルボリューム内のジャーナルデータ領域 820 の内部構造を示す図である。

図 9 は、ジャーナルボリュームの切り替え設定情報を示す図である。

【0036】

ジャーナルグループ管理情報は、ジャーナルグループを管理するための情報であり、各記憶装置システム 120 の制御メモリ 145 に格納される。

【0037】

ここで、「ジャーナルグループ」とは、データボリュームとそのデータのジャーナルを格納するジャーナルボリュームを関連付けたペアをいう。

【0038】

ジャーナルグループ管理情報には、図3に示されるようにジャーナルグループID410、最新ジャーナルシーケンス番号420、データボリューム情報430、ジャーナルボリューム情報440、および、転送グループ情報450が含まれる。

【0039】

ジャーナルグループID410は、ジャーナルグループを一意に決定する識別子である。最新ジャーナルシーケンス番号420は、ジャーナルグループのジャーナルに付与される連続する番号のうち、最も新しく付与された番号である。

【0040】

データボリューム情報430には、ジャーナルグループに含まれる各データボリュームのデータボリューム管理情報431が含まれる。データボリューム管理情報の詳細については後述する。

【0041】

ジャーナルボリューム情報440には、ジャーナルグループに含まれる各ジャーナルボリュームの情報と格納用ジャーナルボリュームID444が含まれる。各ジャーナルボリューム毎の情報には、ボリュームを識別するボリュームID441、ジャーナルボリュームに格納されているジャーナルのうち、最も古いジャーナルのシーケンス番号を示す先頭ジャーナルシーケンス番号442、および、最も新しいジャーナルのシーケンス番号を示す末尾ジャーナルシーケンス番号443が含まれる。

【0042】

格納用ジャーナルボリュームID444は、複数のジャーナルボリュームがジャーナルグループに含まれる場合に、次のジャーナルが格納されるジャーナルボリュームを示すためのものである。プライマリ記憶装置システム120Aにおいてはジャーナル取得処理でジャーナル格納に用いられるジャーナルボリュームであり、セカンダリ記憶装置システム120Bにおいてはジャーナル転送処理でジャーナルの転送先に用いられるジャーナルボリュームを指す。なお、ジャーナルボリュームが1つの場合には格納用ジャーナルボリュームID444には格納用かつ転送用に使用されるジャーナルボリュームのIDが設定される。

【0043】

転送グループ情報450には、ペアとなるジャーナルグループの識別子であるジャーナルグループID451、ジャーナル転送処理222によりどのジャーナルまでセカンダリ記憶装置システムに転送が完了したかを示す転送済みジャーナルシーケンス番号452、セカンダリ記憶装置システム120Bにおいてどのジャーナルまでジャーナル復元処理が完了したかを示す復元済みジャーナルシーケンス番号453、および、転送グループの状態を示すペア状態情報454が含まれる。なお、転送済みジャーナルシーケンス番号452の次のシーケンス番号を有するジャーナルが一番古いジャーナルとなる。セカンダリ記憶装置システム（第二および第四の実施形態ではプライマリ記憶装置システム）は、この一番古いジャーナルが含まれるジャーナルボリュームを転送用のジャーナルボリュームとして特定し、使用する。転送用のジャーナルボリュームと格納用のジャーナルボリュームが一致する場合は、この一番古いジャーナルを含むジャーナルボリュームと格納用ジャーナルボリュームID444で示されるジャーナルボリュームとは一致する。一方、切り替え処理の結果またはあらかじめ転送用のジャーナルボリュームと格納用のジャーナルボリュームとを異ならせた場合、一番古いジャーナルを含むジャーナルボリュームと格納用ジャーナルボリュームID444で示されるジャーナルボリュームとは一致しない。これはセカンダリ記憶装置システム120Bでも同様である。

【0044】

ここで、「転送グループ」とは、転送元のジャーナルグループと転送先のジャーナルグ

ループを関連付けてペアとしたものである。

【0045】

復元済みジャーナルシーケンス番号453は、セカンダリ記憶装置システム120Bのジャーナル処理管理プログラム221からプライマリ記憶装置システム120Aのジャーナル処理管理プログラム221へ通知される。

【0046】

ペア状態454には、ジャーナルグループ内の全データボリュームが複製状態にある「PAIR」、ジャーナルグループ内の一つ以上のデータボリュームが差分コピー（差分コピーについては後述）をおこなっている「COPY」がある。また、ジャーナルグループ内の全データボリュームの複製がおこなわれず、サスペンド状態にある状態で、ジャーナルグループ内でのデータ一貫性がある状態「SUSPEND」と、ジャーナルグループ内でのデータ一貫性がない状態「SUSPEND-E」がある。

【0047】

データボリューム管理情報は、データボリュームを管理するための情報であり、図4に示されるようにデータボリュームの記憶装置システム内で識別するためのボリュームID510、および、ペアであるデータボリュームの情報であるペアボリューム情報520が含まれる。

【0048】

ペアボリューム情報520には、記憶装置システムID521、ペアであるデータボリュームを識別するデータボリュームID522、複製状態を示すペア状態523、差分ビットマップ有効フラグ524、および、差分ビットマップ525が含まれる。

【0049】

記憶装置システムID521はペアであるデータボリュームが存在する記憶装置システムの識別子であり、データボリュームID522は、その記憶装置システムの中でのボリューム識別子である。記憶装置システムID521とデータボリュームID522を組み合わせることにより、データボリュームが一意に決まる。

【0050】

ペア状態523には、データボリュームが複製状態にある（ボリューム内でデータ一貫性がある状態）「PAIR」、差分コピーをおこなっている「COPY」、ボリュームやパスの閉塞等によりコピー処理が停止されサスペンド状態にある「SUSPEND」のうちいずれかの状態を指す。

【0051】

差分ビットマップ有効フラグ524は、差分ビットマップの値が有効であるかどうかを示す。差分ビットマップ525は、PVOLとSVOLのデータに差異がある領域を示す情報である。データボリューム領域を複数の領域に分割し、SUSPEND中にデータボリュームに更新があった場合には、更新のあった領域を示すビットをONにする。SUSPEND後に、PVOLとSVOLの各々の差分ビットマップ525のORを取ったビットマップに基づいてビットがONの領域のみをコピーする（差分コピー）ことで、ペアを複製状態に戻すことができる。差分コピーによりコピー転送量を減らすことができる。差分コピーではコピーが完了した領域はビットをOFFにし、全ビットがOFFになると差分コピーが完了したこととなる。また、ペア生成時には差分ビットマップの全てONにして差分コピーをおこなう（形成コピー）ことで、PVOLの全領域をSVOLにコピーすることができる。

【0052】

ジャーナルメタ情報は、データとジャーナルを関連付けるための管理情報であり、図5に示されるようにデータボリューム情報710とジャーナル情報720を含む。

【0053】

データボリューム情報710には、データが更新された時刻を示す更新時刻711、データが更新されるデータボリュームがジャーナルグループ内のどのデータボリュームであるかを示す格納データボリュームのジャーナルグループ内オフセット712、および、デ

ータボリューム上のデータを格納する先頭アドレスを示すデータ格納アドレス 713 を含む。

【0054】

ジャーナル情報 720 には、ジャーナルボリューム上のジャーナルデータを格納する先頭アドレスを示すジャーナルデータ格納アドレス 721、ジャーナルデータのデータ長 722、および、ジャーナル取得の際に割り当てられたジャーナルグループ内でのジャーナルの通番であるジャーナルシーケンス番号 723 を含む。

【0055】

このジャーナルメタ情報により関連付けられるデータボリュームとジャーナルボリュームは、図 6 に示される如くである。

【0056】

一般に、PVOL、SVOL およびジャーナルボリュームは各々予め定められた論理ブロック単位で管理される（例えば 512 KB）。論理ブロックの各々には、論理ブロックアドレス（以下「LBA」と記す）が付与されている。

【0057】

プライマリジャーナルボリューム 267A は、メタデータ領域 810、および、ジャーナルデータ領域 820 を有する。ジャーナルデータ領域 820 には、先に説明したジャーナルデータ 821A、すなわち、ライトコマンドによって PVOL に書き込まれたデータ 831A のコピーが格納される。メタデータ領域 810 には、先に説明したメタデータ 811A が格納される。メタデータには、更新データのデータ格納アドレス 812A、および、ジャーナルデータの格納アドレス 813A が含まれる。

【0058】

セカンダリジャーナルボリューム 267B も、プライマリジャーナルボリューム 267A と同様にメタデータ領域 810、および、ジャーナルデータ領域 820 を有する。メタデータ領域 810 には、プライマリジャーナルボリュームのメタデータ領域から転送されたメタデータ 811B が格納される。ジャーナルデータ領域 820 には、プライマリジャーナルボリューム 267A のジャーナルデータ領域から転送されたジャーナルデータ 821B（メタデータに対応する）が格納される。

【0059】

メタデータ 811B は PVOL でおこなわれたデータ更新の情報を持ち、そのアドレス情報 813B は対応するジャーナルデータ 821B のアドレスを示す。さらに、ジャーナルデータ 821B を、セカンダリジャーナルボリューム 267B のジャーナルデータ領域 820 から、アドレス 812B に対応する SVOL 266 のアドレスへコピーすることによって、PVOL 265 での更新を SVOL 266 へ反映することができる。

【0060】

各アドレスは LBA により、データ長は論理ブロック数により、各々表わすことができる。また、データが格納されている場所は、データが格納された領域（ジャーナルデータ領域またはメタデータ領域）のベースアドレス（先頭 LBA）との差分（オフセット）で表されても良い。本実施形態において、メタデータのデータ長は一定（例えば 64 バイト）であるが、ジャーナルデータのデータ長は、ライトコマンドで更新されるデータに依存するので一定ではない。

【0061】

ジャーナルグループ定義時に、各記憶装置システム 120 は、設定されるジャーナルボリューム 267 に対して、メタデータ領域 810 およびジャーナルデータ領域 820 の設定をおこなう。具体的には、各領域の先頭 LBA およびブロック数が設定される。

【0062】

次に、プライマリジャーナルボリューム内のジャーナルデータ領域 820 の内部構造について説明する。

【0063】

プライマリジャーナルボリュームが有するジャーナルデータ領域 820 は、図 7 に示さ

れるようにジャーナルデータが格納されているジャーナル格納済み領域 910 と、ジャーナルデータが格納されていない、あるいは、パージ可能なジャーナルデータが格納されているパージ済み領域 920 とに区別される。パージ済み領域 910 は、ジャーナルデータをセカンダリサイトに転送したために、そこに格納されているジャーナルデータを開放しても良いことになった領域であり、PVOL625 の新たなジャーナルデータの格納に使用することができる。

【0064】

図7では、ジャーナルデータ領域 900A とジャーナルデータ領域 900B は、別個の論理ボリュームに格納されていることを示している。

【0065】

ジャーナルデータ領域 900A には、ボリューム内先頭ジャーナルシーケンス番号 931A からボリューム内末尾ジャーナルシーケンス番号 932A までのジャーナルが、ジャーナルデータ領域 900B にはボリューム内先頭ジャーナルシーケンス番号 931B からボリューム内末尾ジャーナルシーケンス番号 932B までのジャーナルが格納されている。ジャーナルを格納するときには、ボリューム内先頭に書き込まれていく。

【0066】

ジャーナルボリュームはサイクリックバッファと同じように繰り返し使用される。つまりジャーナルの末尾の論理ブロックまで使用すると、先頭の論理ブロックが再度使用される。ただし、ジャーナルグループに複数のジャーナルボリュームが含まれる場合は、ジャーナルの末尾の論理ブロックまで使用すると、次のジャーナルボリュームの先頭論理ブロックが使用される。最後のジャーナルボリュームの末尾論理ブロックまで使用されると最初のジャーナルボリュームの先頭論理ブロックに戻る。なお、ジャーナルボリュームの末尾論理ブロックまで使用する前に、途中の論理ブロックから次のジャーナルボリュームの先頭論理ブロックに移っても良い。ジャーナル格納先を次のジャーナルボリュームに移すことを「ジャーナルボリューム切り替え処理」という。

【0067】

図7では、ボリューム内先頭ジャーナルシーケンス番号 931A までジャーナル領域 900A を使用した後、ジャーナルボリュームが切り替わり、ジャーナルデータ領域 900B の先頭に移ったことになる。このためボリューム内先頭ジャーナルシーケンス番号 931A とボリューム内末尾ジャーナルシーケンス番号 932B は連続していることになる。なお、このジャーナルボリューム切り替え処理については、後に詳述する。

【0068】

格納済みジャーナルシーケンス番号 911 は、最も新しいジャーナルを示す。次に取得されるジャーナルは、格納済みジャーナルシーケンス番号 911 に 1 を足した値のジャーナルシーケンス番号が付与され、ジャーナルデータ領域 900B のパージ領域に格納される。ここでは格納用ジャーナルボリュームはジャーナルデータ領域 900B を有するジャーナルボリュームをいう。転送済みジャーナルシーケンス番号 921 に 1 を足したシーケンス番号のジャーナルが最も古いジャーナルである。転送済みジャーナルシーケンス番号 921 と格納済みジャーナルシーケンス番号 911 が等しい場合には、ジャーナルが空であることを意味する。

【0069】

次に、セカンダリジャーナルボリューム内のジャーナルデータ領域 820 の内部構造について説明する。

【0070】

セカンダリジャーナルボリュームが有するジャーナルデータ領域 820 は、図8に示されるように既にSVOL626へのジャーナル復元に使用されたジャーナルデータが格納されている（あるいは、ジャーナルデータが格納されていない）パージ済み領域 1030、SVOL626へのジャーナル復元の対象として指定されたジャーナルデータが格納されている復元中領域 1020、ジャーナル復元の対象となっておらず、プライマリジャーナルボリュームからジャーナルの転送が終わったジャーナルデータが格納されている転送

済み領域 1011、および、プライマリジャーナルボリュームから転送中のジャーナルデータが格納される転送中領域 1010 とに区別される。

【0071】

ジャーナルデータ領域 1000A には、ボリューム内先頭ジャーナルシーケンス番号 931C からボリューム内末尾ジャーナルシーケンス番号 932C までのジャーナルが、ジャーナルデータ領域 1000B にはボリューム内先頭ジャーナルシーケンス番号 931D からボリューム内末尾ジャーナルシーケンス番号 932D までのジャーナルが格納されている。ここではボリューム内先頭ジャーナルシーケンス番号 931C を格納した後にジャーナルボリューム切り替えが起こったことになり、ボリューム内先頭ジャーナルシーケンス番号 931C とボリューム内末尾ジャーナルシーケンス番号 932D は連続する。

【0072】

転送予定ジャーナルシーケンス番号 1012 は、プライマリジャーナルボリュームから転送中のジャーナルの先頭ジャーナルシーケンス番号を示す。ジャーナルは転送予定ジャーナルシーケンス番号 1012 以降のジャーナルが次に転送され、ジャーナルデータ領域 1000B のバージ領域に格納される。ここでは、図 3 の格納用ジャーナルボリューム 444 に格納されるボリューム ID は、転送先のボリュームであるジャーナルデータ領域 1000B を有するジャーナルボリュームを指す。転送済みジャーナルシーケンス番号 1013 は、最後に転送処理が完了したジャーナルのシーケンス番号を指す。

【0073】

復元予定ジャーナルシーケンス番号 1021 は、ジャーナル復元処理の対象とされたジャーナルの先頭シーケンス番号を指す。また、復元済みジャーナルシーケンス番号 1022 より最後に復元処理が完了したジャーナルのシーケンス番号を指す。

【0074】

次に、ジャーナルボリューム切り替え設定情報について説明する。

【0075】

ジャーナルボリューム切り替え設定情報は、ジャーナルボリュームを切り替えるための管理情報である。

【0076】

本発明の処理では、プライマリサイトとセカンダリサイトの両方でジャーナルボリュームを切り替えることが可能である。

【0077】

プライマリサイト 100A のジャーナルボリュームを切り替えは、転送用に用いている（転送元）論理ボリュームと現時点でのジャーナルの格納に用いている論理ボリュームの負荷を分散するために、現時点でのジャーナルの格納に用いている論理ボリュームを切り替える。

【0078】

また、セカンダリサイト 100B のジャーナルボリュームを切り替えは、転送用に用いている（転送先）論理ボリュームと現時点でのジャーナルによるデータ復元に用いている論理ボリュームの負荷を分散するために、転送用に用いている論理ボリュームを切り替える。

【0079】

ジャーナルボリューム切り替え設定情報は、図 9 に示されるように切り替え判定有効フラグ 1110 と、判定情報 1120 と、条件外時動作 1130 よりなる。

【0080】

切り替え判定有効フラグ 1110 は、切り替え判定処理をおこなうか否かのフラグであり、ジャーナルボリューム切り替え判定をおこなわない場合には、判定有効フラグ 1110 は OFF になる。

【0081】

プライマリサイト 100A での切り替え判定処理は、任意のタイミングでおこなうことができるが、あるボリュームのジャーナルの転送が終わって、次に転送すべきジャーナル

の格納されているジャーナルボリュームが、現時点でのジャーナルを格納しているボリュームに使われているときに、切り替え判定処理がおこなうことが有効的である。また、セカンダリサイト 100B での切り替え判定処理も、任意のタイミングでおこなうことができるが、あるボリュームのジャーナルによるデータ復元が終わって、次に復元に用いられるジャーナルの格納されているジャーナルボリュームが、転送先のボリュームに使われているときに、切り替え判定処理がおこなうことが有効的である。ここで、ジャーナルデータは、図 7 および図 8 に用いられるように、シーケンシャルな構造を持つことに注意する。なお、このような切り替え判定処理は、後に詳説する。

【0082】

判定情報 1120 には、未転送ジャーナル量に関する設定値と、未転送ジャーナルのうち最古のジャーナルが更新された時刻と判定時の時間差に関する設定値が含まれる。各々の設定値には、その判定基準を有効にするか否かの有効フラグ 1121、1123、および、格納用ボリューム切り替えするための閾値条件 1122、1124 が含まれる。

【0083】

例えば、プライマリジャーナルボリュームの場合には、未転送ジャーナル量が少ない場合には、現時点でのジャーナルの格納に用いている論理ボリュームと転送に用いている論理ボリュームが同一の論理ボリュームとなってもジャーナル処理の負荷はあまり問題にならないと考えられる。また、セカンダリジャーナルボリュームの場合には、復元予定のジャーナル量が少ない場合には、現時点でのジャーナルによるデータ復元に用いている論理ボリュームと転送に用いている論理ボリュームが同一の論理ボリュームとなってもジャーナル処理の負荷はあまり問題にならないと考えられる。

【0084】

したがって、プライマリジャーナルボリュームの場合には、閾値条件 1122 として、未転送ジャーナル量（格納済みジャーナルシーケンス番号 911 と転送済みジャーナルシーケンス番号 921 の差）が設定閾値以上であること、閾値条件 1124 として、未転送ジャーナル時間差（転送済みジャーナルシーケンス番号 921 に 1 加えたシーケンス番号のジャーナルの更新時刻と判定時の時刻の差）が設定閾値以上であることなどが考えられる。また、セカンダリジャーナルボリュームの場合には、閾値条件 1122 として、未復元ジャーナル量（転送済みジャーナルシーケンス番号 1013 と復元済みジャーナルシーケンス番号 1022 の差）が設定閾値以上であることや、閾値条件 1124 として、未復元ジャーナル時間差（復元済みジャーナルシーケンス番号 1022 に 1 加えたシーケンス番号のジャーナルの更新時刻と判定時の時刻の差）が設定閾値以上であることがある。

【0085】

このような場合には、プライマリジャーナルボリュームの場合では、格納用ジャーナルボリューム切り替えをおこない、セカンダリジャーナルボリュームでは、転送用ボリュームの切り替えをおこなう。

【0086】

条件外動作 1130 には、条件にあわないときのシステム動作を記述する。条件に合わない場合の動作としては、プライマリジャーナルボリュームの場合では、（1）ジャーナル転送を中断する、あるいは、（2）ジャーナル格納に用いているジャーナルボリュームを用いてジャーナル転送をおこない、また、セカンダリジャーナルボリュームの場合では、（1）ジャーナル復元を中断する、（2）転送先に用いているジャーナルボリュームを用いてジャーナル復元をおこなう。

【0087】

プライマリサイト 100A では、ジャーナル切り替え設定情報は、ジャーナルグループ設定時等にホスト 110、管理端末 130、あるいは、保守端末 122 の GUI を介してユーザによって設定され、プライマリ記憶装置システム 120A の制御メモリ 145 に格納される。セカンダリサイト 100B でも、ジャーナル切り替え設定情報は、同様に設定され、セカンダリ記憶装置システム 120B の制御メモリ 145 に格納される。

(III) データ処理システムの処理概要

(III-1) データ処理システムの処理概要

先ず、図10を用いて本発明の第一の実施形態に係るデータ処理システムの処理概要について説明する。

【0088】

図10は、第一の実施形態に係るデータ処理システムの概略処理を示すフローチャートである。

【0089】

先ず、ユーザは、ホスト110、管理端末130、あるいは、保守端末122が持つGUI (Graphical User Interface) 等を用いて記憶装置システムにペア生成コマンドを入力する(ステップ301)。

【0090】

ペア生成コマンドは、データ複製のコピー元となるプライマリ記憶装置システム120Aが有するPVOL625と、PVOLに格納されるデータの複製先となるセカンダリ記憶装置システム120Bが有するSVOL626とを、ペアとして関連付けるコマンドである。

【0091】

次に、各々のサイトのジャーナル処理管理プログラム221を用いて、プライマリサイト100Aにおいて、記憶装置システム120Aで指定されたPVOLに対応するジャーナルを格納するボリュームを割り当てるように制御し、また、セカンダリサイト100Bにおいて、セカンダリ記憶装置システム120Bで指定されたSVOL626に対応するジャーナルを格納するジャーナルボリュームを割り当てるように制御する(ステップ302、ステップ303)。

【0092】

プライマリサイト100Aでは、PVOL625とそれに割り当てられたジャーナルボリューム627Aにより、ジャーナルグループが形成され、セカンダリサイト100Bでは、SVOL626とそれに割り当てられたジャーナルボリューム627Bにより、ジャーナルグループが形成される。ジャーナルグループを形成する際に、ジャーナルボリュームに複数のボリュームの集合を割り当てることができる。

【0093】

ペア生成コマンドは、PVOLのジャーナルグループとSVOL626のジャーナルグループも関連付け、転送グループとする(ステップ304)。ジャーナルグループの設定時に、ジャーナルボリューム切り替え情報を設定しても良い。ジャーナルボリューム切り替え設定の詳細は後述する。

【0094】

なお、ジャーナルグループ形成の際には、単一のデータボリュームに限らず複数のデータボリュームの集合を割り当てることができる。このPVOL集合におけるデータの更新と同様にSVOL集合でもデータが更新されるため、このデータボリューム集合内でデータの一貫性が保持される。

【0095】

次に、転送グループが作成された後、ジャーナル処理が実行される(ステップ305)。ジャーナル処理は、ジャーナル取得処理、ジャーナル転送処理、ジャーナル復元処理の各処理をいう。ジャーナル取得処理を開始は、ユーザからのジャーナルの取得を指示するコマンド(以下「ジャーナル取得開始コマンド」)をプライマリ記憶装置システムが受信したことを契機に、プライマリ記憶装置システム120でおこなわれる。ジャーナル処理の詳細については後述する。

【0096】

一方、ジャーナル取得処理が開始される前にPVOL625に格納されていたデータは、ジャーナル転送処理が開始されてもセカンダリ記憶装置システムには転送されない。別途PVOL625からSVOL626へこれらのデータ(以下「イニシャルデータ」)をコピーする必要がある。イニシャルデータをコピーする処理を「形成コピー」という。本

実施形態においては、イニシャルデータを P V O L 6 2 5 から S V O L 6 2 6 に転送する形成コピー処理が実行される（ステップ 3 0 6）。イニシャルデータは、P V O L 6 2 5 のボリューム先頭領域から末尾まで転送される。

【0097】

(III-2) ジャーナル処理の詳細

次に、図 1 1 を用いてジャーナル処理の詳細について説明する。

図 1 1 は、本発明の第一の実施形態に係るジャーナル処理の動作を示す図である。

【0098】

記憶装置システム 1 2 0 A および 1 2 0 B により、それぞれジャーナル処理管理プログラム 2 2 1 を実行してジャーナル処理を制御する。

【0099】

プライマリ記憶装置システム 1 2 0 A は、ジャーナル取得・復元プログラム 2 2 3 の内、ジャーナル取得プログラム 6 2 3 を実行する。ジャーナル取得プログラム 6 2 3 を実行することによって、プライマリ記憶装置システム 1 2 0 A は、P V O L 6 2 5 に書き込まれるデータのコピーをジャーナルデータとして、ジャーナルボリューム 6 2 7 A に格納する。また、プライマリ記憶装置システム 1 2 0 A はジャーナルの一部としてメタデータもジャーナルボリューム 6 2 7 A に格納する。これらの処理がジャーナル取得処理である。

【0100】

一方、セカンダリ記憶装置システム 1 2 0 B は、ジャーナル取得・復元プログラム 2 2 3 の内、ジャーナル復元プログラム 6 2 4 を実行してジャーナル復元処理をおこなう。ジャーナル復元プログラム 6 2 4 は、ジャーナルボリューム 6 2 7 B に格納されたジャーナルに基づきデータを復元して、P V O L 6 2 5 で更新されたデータを S V O L 6 2 6 に反映する。

【0101】

これらのような処理を、図 2 に示したシステムで実行する場合には、図 6 に示されるようになる。

【0102】

プライマリ記憶装置システム 1 2 0 A で P V O L 6 2 5 のジャーナル取得処理が開始されると、プライマリ記憶装置システム 1 2 0 A は、プライマリホスト 1 1 0 A から P V O L 6 2 5 への書き込み処理（矢印 6 0 1）に応じて、ジャーナルを作成し、作成したジャーナルをジャーナルボリューム 6 2 7 A に格納する（矢印 6 0 2）。ここでジャーナル取得プログラムは、プライマリ記憶装置システム 1 2 0 A の制御メモリ 1 4 5 上に格納されているジャーナルグループ管理情報から、最新ジャーナルシーケンス番号や格納用ジャーナルボリューム I D 等の情報を取得し、ジャーナル格納先を決定し、かつ、メタデータ情報を作成する。

【0103】

セカンダリ記憶装置 1 2 0 B はジャーナル処理管理プログラム 2 2 2 を実行してプライマリ記憶装置 1 2 0 A から、ジャーナル作成状況に関する情報（例えば、ジャーナルボリューム内のジャーナルの容量、ジャーナルの一番古い時刻等）を取得する（矢印 6 0 3）。

【0104】

セカンダリ記憶装置システム 1 2 0 B は、ジャーナル処理管理プログラムを実行し、G U I を介したユーザからの指示の入力あるいは予め定められていたスケジュール（例えば、プライマリ記憶装置システム 1 2 0 A で一定量のジャーナルがジャーナルボリュームに格納されたとき、または、一定期間ごと等）に従い、セカンダリ記憶装置システム 1 2 0 B のジャーナル転送プログラム 2 2 2 に対し、ジャーナル転送要求を出す（矢印 6 0 4）。

【0105】

ジャーナル転送要求には、コピーすべきジャーナル（複数でも良い）、そのジャーナルが格納されているジャーナルボリューム、そのジャーナルボリュームを有する記憶装置シ

システム 120（ここではプライマリ記憶装置システム 120A）を指定する情報、および、コピーしたジャーナルが格納されるジャーナルボリュームを指定する情報が含まれている。これらの情報はジャーナル処理管理プログラム 221 が制御メモリ 145 上のジャーナルグループ管理情報から取得した情報から作成される。

【0106】

ジャーナル転送要求を受けたジャーナル転送プログラム 222 は、リードコマンドをプライマリ記憶装置システム 120A に対して発行する（矢印 605）。このリードコマンドを受けたプライマリ記憶装置システム 120A は、リードコマンドで指定されたジャーナルをセカンダリ記憶装置システム 120B に送信する（矢印 606）。

【0107】

セカンダリ記憶装置システム 120B に送信されたジャーナルが格納されていた、プライマリ記憶装置システム 120A のジャーナルボリュームの領域はパージ（開放）され、新たなジャーナルの格納に利用することができる。なお、パージは送信後すぐにおこなう必要はない。パージを定期的におこなっても良いし、ユーザからの指示に従ってパージしても良い。

【0108】

ジャーナルを受信したセカンダリ記憶装置システム 120B は、ジャーナル転送要求で指定されたジャーナルボリューム 627B に、受信したジャーナルを格納する。

【0109】

その後、セカンダリ記憶装置システム 120B のジャーナル処理管理プログラム 221 は、セカンダリ記憶装置システム 120B のジャーナル復元プログラム 624 に対して、ジャーナル復元要求を発行する（矢印 607）。ジャーナル復元要求を受けたジャーナル復元プログラム 624 は、ジャーナルボリューム 627B のジャーナルに基づき、SVOL 626 にデータの復元をおこなう（矢印 608）。なお、復元に用いられたジャーナルが格納されていた領域はパージされ、新たなジャーナルの格納に利用することができる。

【0110】

（III-3）ジャーナルボリューム切り替え処理

次に、図 12 ないし図 14 を用いてジャーナルボリューム切り替え処理について説明する。

図 12 は、ジャーナルボリュームの切り替えを説明する概念図である。

図 13 は、プライマリジャーナルボリュームの切り替え処理を示すフローチャートである。

図 14 は、セカンダリジャーナルボリュームの切り替え処理を示すフローチャートである。

【0111】

既に述べてきているように、本発明では、プライマリサイト 100A からセカンダリサイト 100B にジャーナルを転送するために、ジャーナルの格納、ジャーナル復元と転送に用いている論理ボリュームが同一となることを回避してシステムの負荷を軽減するために、ジャーナルボリュームの切り替え処理をおこなう。

【0112】

今、図 12 に示されるように、プライマリサイト 100A には、論理ボリューム PA、論理ボリューム PB、論理ボリューム PC があり、論理ボリューム PB がジャーナル格納用として使用され、論理ボリューム PA が転送元の論理ボリュームとして転送用に使用されているとする。

【0113】

ジャーナルの順序は、ボリュームの上の方が古く、論理ボリューム PA のジャーナルの最後が、論理ボリューム PB のジャーナルの最初につながっているものとする。

【0114】

さて、論理ボリューム PA 内のジャーナルが転送され終わると、論理ボリューム PB のジャーナルを転送することになるが、このときに、ジャーナル格納用の論理ボリュームを

、論理ボリューム P B から論理ボリューム P C に切り替える。

【0115】

また、セカンダリサイト 110 B には、論理ボリューム S A、論理ボリューム S B、論理ボリューム S C があり、論理ボリューム S C がジャーナル復元用として使用され、論理ボリューム S A が転送先の論理ボリュームとして転送用に使用されているとする。

【0116】

ジャーナルの順序は、ボリュームの上の方が古く、論理ボリューム S C のジャーナルの最後が、論理ボリューム S A のジャーナルの最初につながっているものとする。

【0117】

このときには、論理ボリューム S C 内のジャーナルが復元され終わると、論理ボリューム S A のジャーナルを復元することになるが、このときに、ジャーナル転送先の論理ボリュームを、論理ボリューム S A から論理ボリューム S B に切り替える。

【0118】

上記では、ジャーナル転送用の論理ボリュームとジャーナル復元用の論理ボリュームにあるジャーナルが全て転送されたり、復元されたりしたタイミングで切り替える例を示したが、ジャーナルボリューム切り替えは、ユーザによって入力された切り替えコマンドに従っておこなわれても良い。また、一定期間毎や、設定時刻におこなわれても良い。また、予めホスト 110、管理端末 130、あるいは、保守端末 122 のインタフェースを介して設定された条件を満たした場合におこなわれても良い。

【0119】

また、プライマリサイト 100 A において、図 11 に示した矢印 605 のジャーナル転送を要求するコマンドを受け付けたタイミングで切り替えることにしてもよい。セカンダリサイト 100 B において、ジャーナル転送が開始されたタイミングで切り替えることにしてもよい。

【0120】

このようなジャーナルボリュームの切り替えは、具体的には、プライマリサイト 100 A では、複数のプライマリジャーナルボリュームがある場合に、図 7 に示した格納済みジャーナルシーケンス番号 911 と転送済みジャーナルシーケンス番号 921 の差を制御することによって、ジャーナル取得処理に用いるジャーナルボリュームとジャーナル転送処理に用いるジャーナルボリュームを異なるボリュームにすることが可能である。

【0121】

また、セカンダリサイト 100 B では、複数のセカンダリジャーナルボリュームがある場合に、図 8 に示した復元済みジャーナルシーケンス番号 1022 と転送予定ジャーナルシーケンス番号 1012 の差を制御することによって、ジャーナル復元に用いるジャーナルボリュームとジャーナル転送処理に用いるジャーナルボリュームを異なるボリュームにすることが可能である。

【0122】

なお、プライマリサイト 100 A で、ジャーナル転送処理対象となるジャーナル（以下「未転送ジャーナル」）がジャーナル取得処理に用いられているジャーナルボリューム（格納用ジャーナルボリューム）以外のボリュームにない場合（すなわち、格納用ジャーナルボリュームのボリューム内末尾ジャーナルシーケンス番号 932 B が、転送済みジャーナルシーケンス番号 921 に 1 加えた番号である場合）、格納用ジャーナルボリュームの切り替えをおこなわなければ、ジャーナル取得とジャーナル転送が同一ジャーナルボリュームに対しておこなわれることになる。

【0123】

次に、図 13 のフローチャートを追いながら、プライマリジャーナルボリューム切り替え処理について説明する。

【0124】

ジャーナル転送処理をおこなう際に、ジャーナルボリューム切り替え判定フラグをチェックする（ステップ 1201）。

【0125】

切り替え判定が有効でなければジャーナル転送処理をおこなう（ステップ1208）。切り替え判定が有効であれば、格納用ジャーナルボリューム以外のジャーナルボリュームに未転送ジャーナルがあるか否かをチェックし（ステップ1202）、未転送ジャーナルがあれば、そのジャーナルボリュームのジャーナル転送処理をおこなう（ステップ1208）。未転送ジャーナルがなければ、未転送ジャーナル量による判定が有効であり、かつ閾値条件を満たしているか否かをチェックし（ステップ1203）、有効かつ条件合致であれば、格納用ジャーナルボリュームを次のジャーナルボリュームに切り替えて（ステップ1207）、ジャーナル転送をおこなう（ステップ1208）。ジャーナル量判定が無効または条件に合わなければ、未転送ジャーナルの最古更新時刻と現在時刻との時間差による判定が有効であり、かつ、閾値条件を満たすか否かをチェックし（ステップ1204）、有効かつ条件合致であれば格納用ジャーナルボリュームを次のジャーナルボリュームに切り替えて（ステップ1207）、ジャーナル転送をおこなう（ステップ1208）。時間差判定が無効または条件に合わなければ、設定情報に合わせて（ステップ1205）、格納用ジャーナルボリュームを用いてジャーナル転送をおこなうか（ステップ1208）、次のジャーナル転送要求があるまでジャーナル転送を中断する（1206）。ステップ1207で格納用ジャーナルボリュームを切り替える際、プライマリ記憶装置システム120Aの記憶制御装置140は、制御メモリ145に格納されているジャーナルグループ管理情報の格納用ジャーナルボリュームIDを、切り替え先の新たなジャーナルボリュームを示す情報に書き換える。

【0126】

次に、図14のフローチャートを追いながら、セカンダリジャーナルボリューム切り替え処理について説明する。

【0127】

ジャーナル復元処理をおこなう際に、ジャーナルボリューム切り替え判定フラグをチェックする（ステップ1301）。切り替え判定が有効でなければ、ジャーナル復元処理をおこなう（ステップ1308）。切り替え判定が有効であれば、格納用ジャーナルボリューム以外のジャーナルボリュームに未復元ジャーナルがあるか否かをチェックし（ステップ1302）、未復元ジャーナルがあれば、ジャーナル復元処理をおこなう（ステップ1308）。未復元ジャーナルがなければ、未復元ジャーナル量による判定が有効であり、かつ、閾値条件を満たしているか否かをチェックし（ステップ1303）、有効かつ条件合致であれば、転送先ジャーナルボリュームを次のジャーナルボリュームに切り替えて（ステップ1307）、ジャーナル復元をおこなう（ステップ1308）。ステップ1307で転送先ジャーナルボリュームを切り替える際、セカンダリ記憶装置システム120Bの記憶制御装置140は、制御メモリ145に格納されているジャーナルグループ管理情報の格納先ジャーナルボリュームIDを、切り替え先の新たなジャーナルボリュームを示す情報に書き換える。ジャーナル量判定が無効または条件に合わなければ未復元ジャーナルの最古更新時刻と現在時刻との時間差による判定が有効であり、かつ閾値条件を満たすかチェックし（ステップ1304）、有効かつ条件合致であれば転送先ジャーナルボリュームを次のジャーナルボリュームに切り替えて（ステップ1307）、ジャーナル復元をおこなう（ステップ1308）。時間差判定が無効または条件に合わなければ、設定情報に合わせて（ステップ1305）、格納用ジャーナルボリュームを用いてジャーナル復元をおこなうか（ステップ1308）、次のジャーナル復元要求があるまでジャーナル復元を中断する（1306）。

【0128】**〔実施形態2〕**

次に、本発明に係る第二の実施形態を、図15を用いて説明する。

図15は、本発明の第二の実施形態に係るジャーナル処理の動作を示す図である。

【0129】

第一の実施形態のジャーナル処理では、図11に示されるようにジャーナル転送処理に

においてセカンダリ記憶装置システム120Bがプライマリ記憶装置システム120Aにジャーナル転送を要求するリードコマンドを発行して(矢印605)、転送をおこなった。本実施形態では、プライマリ記憶装置システム120Aがセカンダリ記憶装置システム120Bからのリードコマンドを待つのではなく、プライマリ記憶装置システム120Aからセカンダリ記憶装置120Bに対してデータを書き込むライトコマンドを発行することにより、ジャーナル転送処理をおこなう。

【0130】

先ず、プライマリサイト100Aにおいて、PVOLデータの更新(矢印601)に対するジャーナル取得処理は(矢印602)、第一の実施形態と同様である。

【0131】

プライマリ記憶装置システム120Aのジャーナル処理管理プログラム221Aは、ジャーナル転送プログラム222にジャーナル転送要求を発行する(矢印1404)。ジャーナル転送要求には、記憶装置システム120Bへ送信すべきジャーナルが格納されているジャーナルボリューム、記憶装置システム120Bを指定する情報、および、そのジャーナルを記憶装置システム120Bで格納すべきジャーナルボリュームを指定する情報等が含まれる。これらの情報は制御メモリ145上に格納されたジャーナルグループ管理情報から取得する。

【0132】

ジャーナル転送要求を受け取ったジャーナル転送プログラム222は、ライトコマンドをセカンダリ記憶装置システム120Bに発行することで指定されたジャーナルをセカンダリ記憶装置システムに送信する(矢印1406)。セカンダリ記憶装置システム120Bは、プライマリ記憶装置システムからライトコマンドとして受信したジャーナルを、そのコマンドで指定されたセカンダリジャーナルボリュームの領域に格納する。

【0133】

その後、セカンダリサイト100Bにおけるジャーナル復元処理(ステップ608)は、第一の実施形態と同様である。また、本実施形態におけるジャーナルボリューム切り替え処理も、第一の実施形態におけるジャーナルボリューム切り替え処理と同様である。

【0134】

〔実施形態3〕

次に、本発明に係る第三の実施形態を、図16を用いて説明する。

図16は、本発明の第三の実施形態に係るジャーナル処理の動作を示す図である。

【0135】

本実施形態のデータ処理システムは、図16に示されるように、ジャーナル処理管理プログラム1521が、記憶装置システム120に含まれるのではなく、ホスト110内に含まれる点で第一の実施形態と異なる。そして、ジャーナル処理管理プログラム1521がプライマリホスト110Aとセカンダリサイト110Bとを接続する通信線により通信をおこなう。

【0136】

先ず、プライマリサイト100Aにおいて、PVOLデータの更新(矢印601)に対するジャーナル取得処理は(矢印602)、第一の実施形態と同様である。

【0137】

プライマリホスト110Aは、ジャーナル処理管理プログラム1521を実行して特定のコマンド(以下「ジャーナル作成状況取得コマンド」)を発行することで、プライマリ記憶装置システム120Aの制御メモリ145上に格納されているジャーナルボリューム管理情報からジャーナル作成状況に関する情報(例えば、ジャーナルの容量)を取得する(矢印1509)。

【0138】

プライマリホスト110Aが取得したジャーナル作成状況に関する情報は、セカンダリホスト110Bへ通知される(矢印1503)。

【0139】

セカンダリホスト 110B は、ジャーナル処理管理プログラム 1521 を実行して、GUI を介したユーザからの指示の入力あるいは予め定められたスケジュール（例えば、プライマリ記憶装置システム 120A で一定量以上のジャーナルが格納されたとき、または一定期間毎）に従い、セカンダリ記憶装置システム 120B に対して、ジャーナル転送要求を発行する（矢印 1504）。

【0140】

ジャーナル転送要求には、コピーすべきジャーナル、そのジャーナルが格納されているジャーナルボリューム、ジャーナルボリュームを有する記憶装置システム 120 を指定する情報、および、コピーしたジャーナルが格納されるジャーナルボリュームを指定する情報が含まれる。

【0141】

ジャーナル転送要求を受信したセカンダリ記憶装置システム 120B は、ジャーナル転送プログラム 222 を実行することで、リードコマンドをプライマリ記憶装置システム 120A に発行する。リードコマンドを受け取ったプライマリ記憶装置 120A は、リードコマンドで指定されたジャーナルをセカンダリ記憶装置システム 120B に送信する（矢印 606）。セカンダリ記憶装置システム 120B に送信されたジャーナルが格納されていた領域は、パージ（開放）され、新たなジャーナルの格納に利用することができる。

【0142】

ジャーナルを受信したセカンダリ記憶装置システム 120B は、ジャーナル転送要求で指定されたジャーナルボリューム 627B に、受信したジャーナルを格納する。

【0143】

その後、セカンダリホスト 110B は、セカンダリ記憶装置システム 120B に対して、ジャーナル復元要求を発行する（矢印 1507）。

【0144】

ジャーナル復元要求を受信したセカンダリ記憶装置システム 120B は、ジャーナル復元プログラム 624 を実行して、ジャーナルボリューム 627B から SVOL 626 にデータの復元をおこなう（矢印 608）。復元が終わったジャーナルが格納されていた領域はパージされ、新たなジャーナルの格納に利用することができる。

【0145】

本実施形態におけるジャーナルボリューム切り替え処理は、第一の実施形態におけるジャーナルボリューム切り替え処理と同様である。

【0146】

〔実施形態 4〕

次に、本発明に係る第四の実施形態を、図 17 を用いて説明する。

図 17 は、本発明の第四の実施形態に係るジャーナル処理の動作を示す図である。

【0147】

本実施形態のデータ処理システムは、図 17 に示されるようにジャーナル転送処理においてプライマリ記憶装置システム 120A がセカンダリ記憶装置システム 120B からのリードコマンドを待つのではなく、プライマリ記憶装置システム 120A からセカンダリ記憶装置 120B に対してデータを書き込むライトコマンドを発行する点で第一の実施形態と異なる。また、ジャーナル処理管理プログラム 1621 が、記憶装置システム 120 上ではなく、ホスト 110 上で実行される点でも第一の実施形態と異なる。また、セカンダリ記憶装置システム 120B がジャーナル復元処理を実行するのではなく、セカンダリホスト 110B がセカンダリジャーナルボリューム 627B から復元に使用するジャーナルを読み出し、SVOL 626 のデータを復元する点でも第一の実施形態と異なる。本実施形態において、ジャーナル復元プログラムはセカンダリホスト 110B 上で実行されるプログラムである。

【0148】

本実施形態において、ジャーナル転送処理の主体がプライマリ記憶装置システム 120A であること、ジャーナル復元処理をおこなうのがセカンダリホスト 110B であること

から、セカンダリ記憶装置システム 120B には、特殊な機能を有しない一般的な記憶装置を用いることができる。

【0149】

プライマリサイト 100A において、PVOL データの更新（矢印 601）に対するジャーナル取得処理（矢印 602）は第一の実施形態と同様である。

【0150】

プライマリホスト 110A は、ジャーナル処理管理プログラム 1621 を実行してジャーナル作成状況取得コマンドを発行することで、プライマリ記憶装置システム 120A の制御メモリ 145 上に格納されているジャーナルグループ管理情報からジャーナル作成状況に関する情報（例えば、ジャーナルの容量）を取得する（矢印 1609）。

【0151】

プライマリホスト 110A が取得したジャーナル作成状況に関する情報は、セカンダリホスト 110B へ通知される（矢印 1603）。

【0152】

プライマリホスト 110A は、ジャーナル処理管理プログラム 1521 を実行して、GUI を介したユーザからの指示の入力あるいは予め定められたスケジュール（例えば、プライマリ記憶装置システム 120A で一定量以上のジャーナルが格納されたとき、または、一定期間毎）に従い、プライマリ記憶装置システム 120A に対して、ジャーナル転送要求を発行する（矢印 1604）。

【0153】

ジャーナル転送要求には、セカンダリ記憶装置システム 120B へ送信すべきジャーナルが格納されているジャーナルボリューム、セカンダリ記憶装置システム 120B を指定する情報、および、そのジャーナルを指定する情報等が含まれている。

【0154】

ジャーナル転送要求を受け取ったプライマリ記憶装置システム 120A は、ライトコマンドをセカンダリ記憶装置システム 120B に発行することで、指定されたジャーナルをセカンダリ記憶装置システム 120B に送信する（矢印 1606）。

【0155】

セカンダリ記憶装置システム 120B は、プライマリ記憶装置システム 120A からのライトコマンドとして受信したジャーナルを、ライトコマンドで指定されたセカンダリジャーナルボリュームの領域に格納する。

【0156】

セカンダリホスト 110B は、ジャーナル復元プログラム 1624 を実行することで、セカンダリジャーナルボリューム 627B からジャーナルを読み出して SVOL 626 にデータの復元をおこなう（矢印 1608）。

【0157】

セカンダリジャーナルグループの管理は、セカンダリホスト 110B がおこない、ジャーナル転送要求の作成必要な情報（格納用ジャーナルボリュームの情報など）をプライマリホスト 110A に通知する。復元が終わったジャーナルが格納されていた領域はパージされ、新たなジャーナルの格納に利用することができる。

本実施形態におけるジャーナルボリューム切り替え処理は、第一の実施形態におけるジャーナルボリューム切り替え処理と同様である。

【0158】

〔上記の実施形態による本発明のデータ処理システムの特徴〕

上述した本発明のデータ処理システムでは、記憶装置システムがジャーナル取得・復元および転送処理をおこない、ジャーナル管理、コピー状態管理は、ホストあるいは記憶装置システムがおこなう構成とした。これによりプライマリサイトとセカンダリサイト間でのデータ複製の実際のデータ転送は、記憶装置システム間のファイバケーブル等で実施される。このことによりホスト間の一般回線のトラフィックを最小限に抑え、またデータの転送は高速な回線となりコピー処理性能の向上が可能である。

【0159】

さらに、記憶装置システムがジャーナルをライトコマンドにより他の記憶装置システムに書き出す機能を有し、そのジャーナルをホストが読み出し、ジャーナル復元することで、セカンダリサイトの記憶装置システムに特別な機能を持たせなくてもデータ複製を実現することが可能となる。

【0160】

さらに、ジャーナルグループは複数のジャーナルボリュームを有し、格納用ジャーナルボリュームを切り替え、プライマリサイトにおいてジャーナル取得処理とジャーナル転送処理を、セカンダリサイトにおいてジャーナル転送処理とジャーナル復元処理を異なるジャーナルボリュームに対しておこなう。これによりジャーナルボリュームへのアクセス負荷を分散することができる。このことによりボリュームへの負荷集中によるボリュームリード処理やライト処理の遅延が軽減され、システムトータル性能の向上が可能である。

【図面の簡単な説明】

【0161】

【図1】本発明の第一の実施形態に係るデータ処理システムのハードウェア構成を示す図である。

【図2】本発明の第一の実施形態に係るデータ処理システムの機能構成を示す図である。

【図3】ジャーナルグループ管理情報を示す図である。

【図4】データボリューム管理情報を示す図である。

【図5】ジャーナルメタ情報を示す図である。

【図6】データボリュームとジャーナルボリュームの対応を示す図である。

【図7】プライマリジャーナルボリューム内のジャーナルデータ領域820の内部構造を示す図である。

【図8】セカンダリジャーナルボリューム内のジャーナルデータ領域820の内部構造を示す図である。

【図9】ジャーナルボリュームの切り替え設定情報を示す図である。

【図10】第一の実施形態に係るデータ処理システムの概略処理を示すフローチャートである。

【図11】本発明の第一の実施形態に係るジャーナル処理の動作を示す図である。

【図12】ジャーナルボリュームの切り替えを説明する概念図である。

【図13】プライマリジャーナルボリュームの切り替え処理を示すフローチャートである。

【図14】セカンダリジャーナルボリュームの切り替え処理を示すフローチャートである。

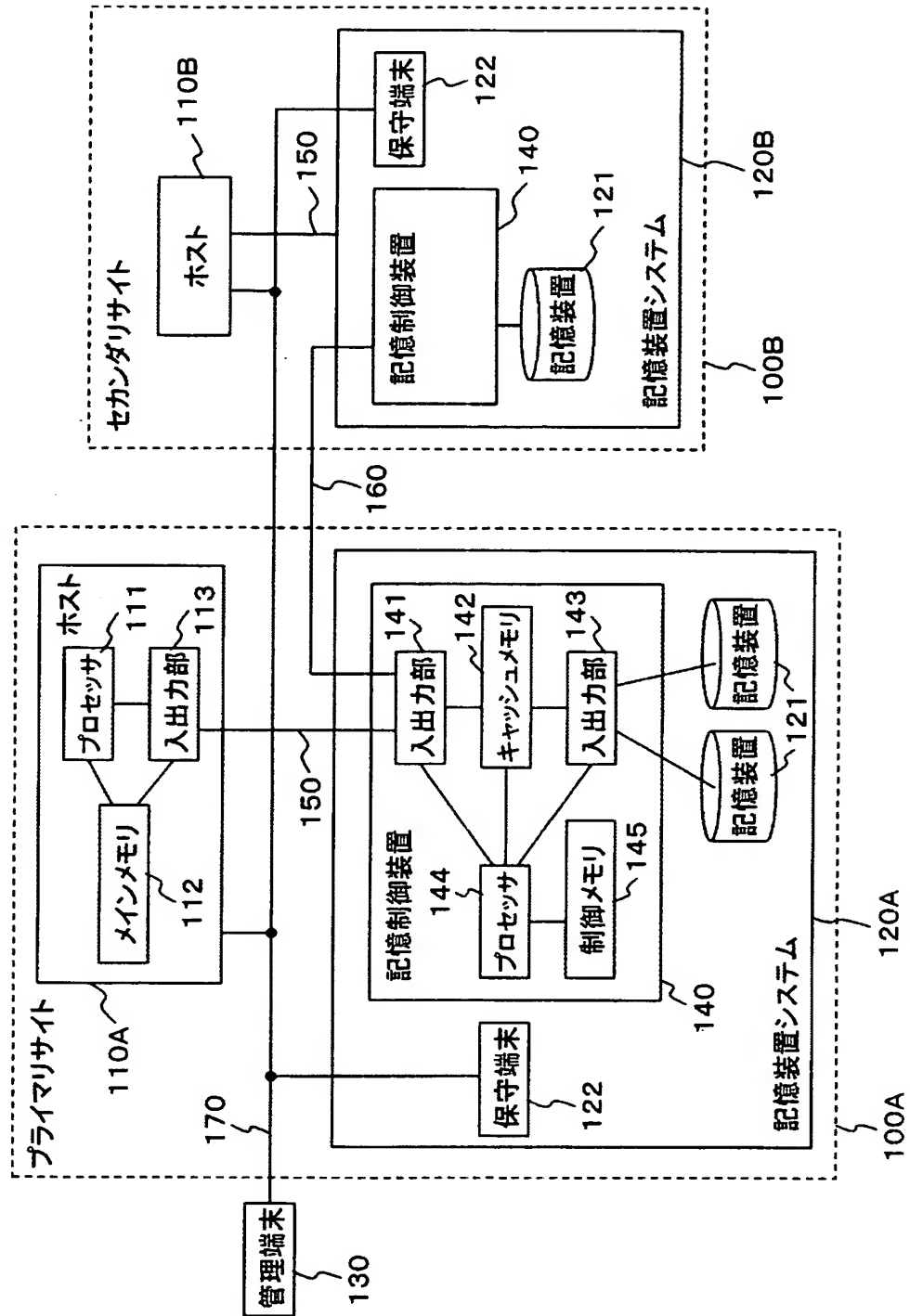
【図15】本発明の第二の実施形態に係るジャーナル処理の動作を示す図である。

【図16】本発明の第三の実施形態に係るジャーナル処理の動作を示す図である。

【図17】本発明の第四の実施形態に係るジャーナル処理の動作を示す図である。

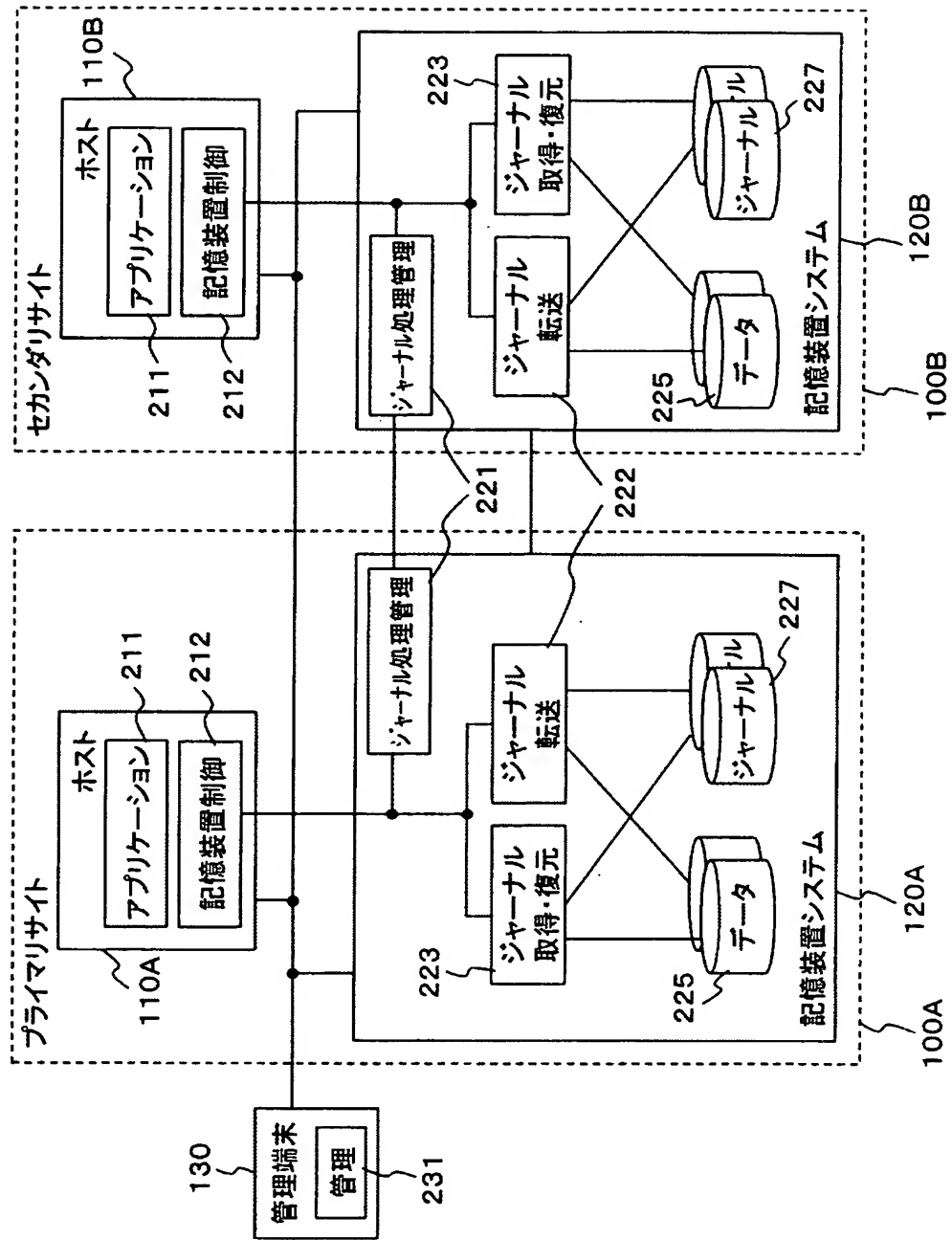
【書類名】 図面
【図 1】

図 1



【図2】

図 2



【図 3】

図 3

410	ジャーナルグループID			
420	最新ジャーナルシーケンス番号			
430	データボリューム情報	データボリューム1	データボリューム管理情報	431
		データボリューム2	...	
		...		
440	ジャーナルボリューム情報	ジャーナルボリューム1	ボリュームID	441
			先頭ジャーナルシーケンス番号	442
			末尾ジャーナルシーケンス番号	443
		ジャーナルボリューム2	...	
450	転送グループ情報	...		
		格納用ジャーナルボリュームID		444
		ジャーナルグループID		451
		転送済みジャーナルシーケンス番号		452
		復元済みジャーナルシーケンス番号		453
		ペア状態		454

【図 4】

図 4

510	ボリュームID		
520	ペアボリューム情報	記憶装置システムID	521
		データボリュームID	522
		ペア状態	523
		差分ビットマップ有効フラグ	524
		差分ビットマップ	525

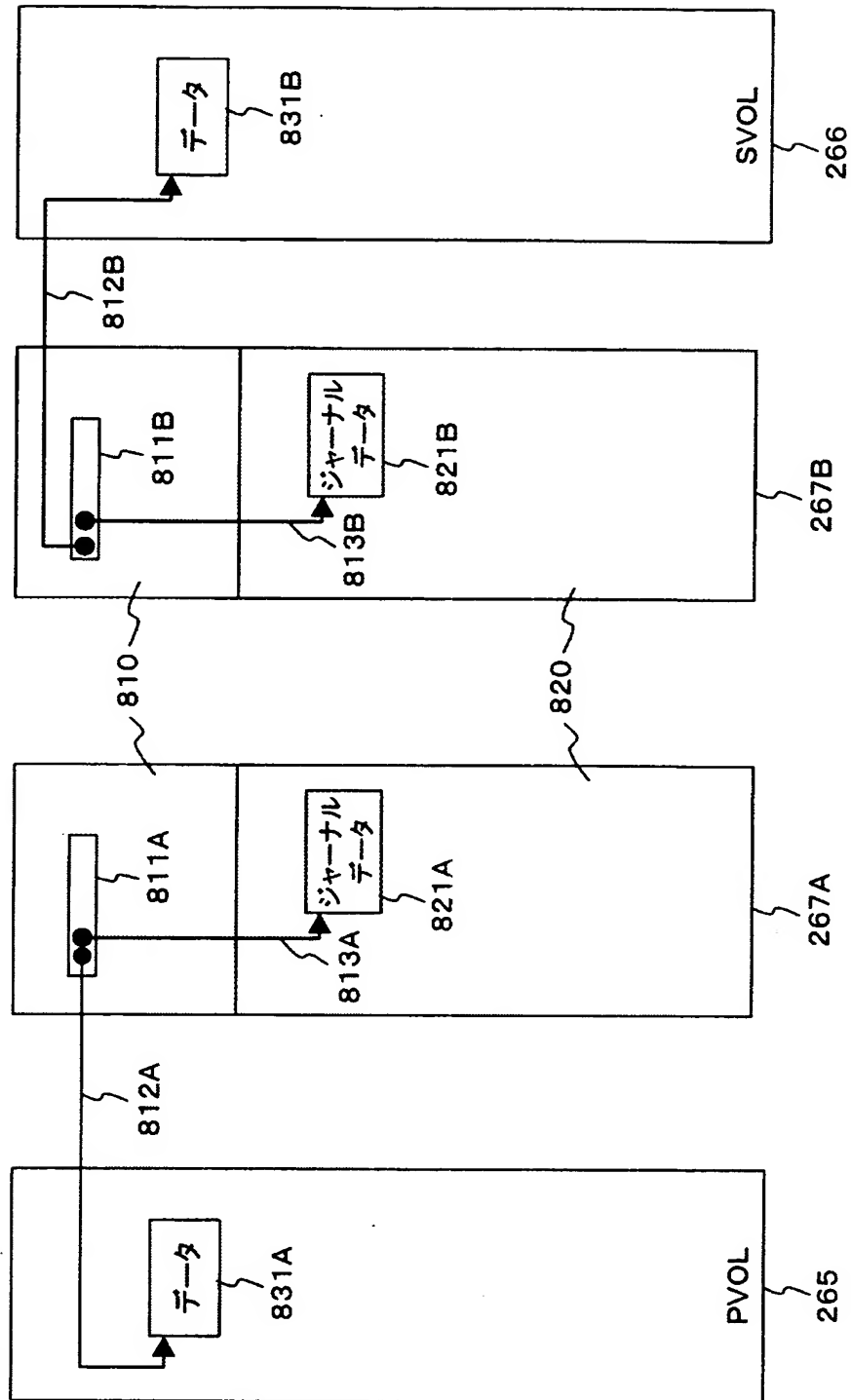
【図 5】

図 5

710	データボリューム情報	更新時刻	711
		格納データボリュームのジャーナルグループ内オフセット	712
		データ格納アドレス	713
720	ジャーナル情報	ジャーナルデータ格納アドレス	721
		データ長	722
		ジャーナルシーケンス番号	723

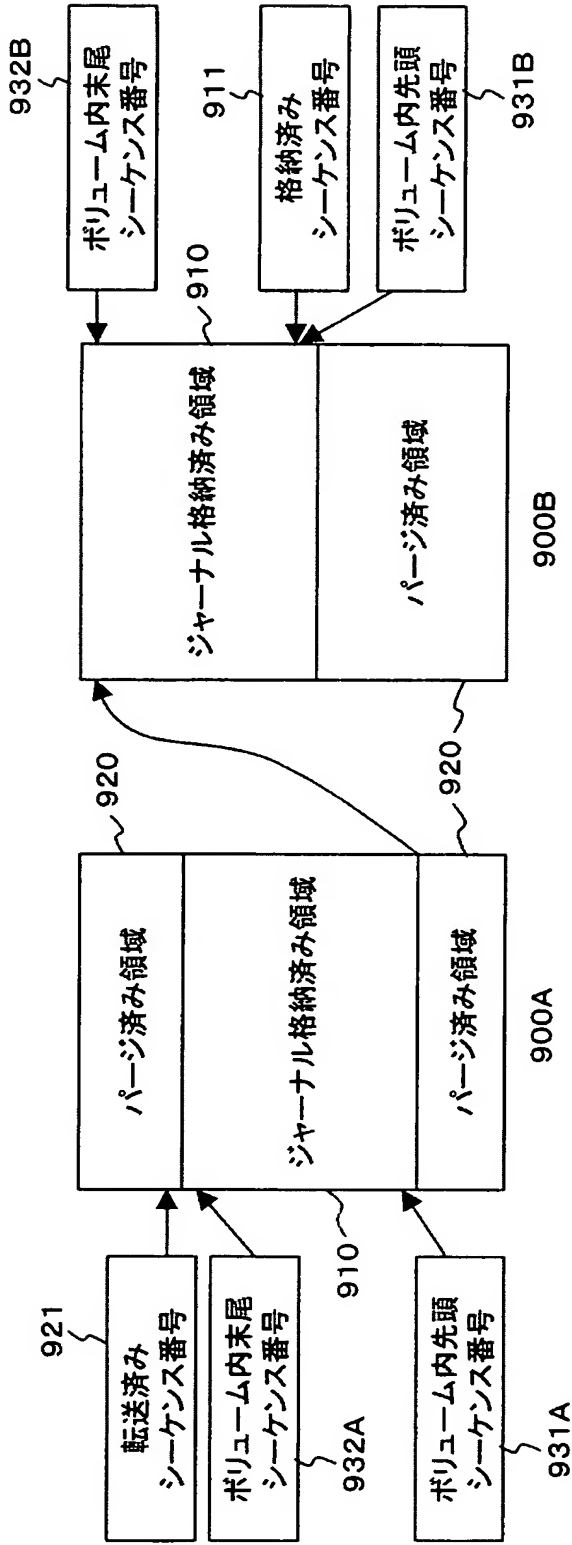
【図 6】

図 6



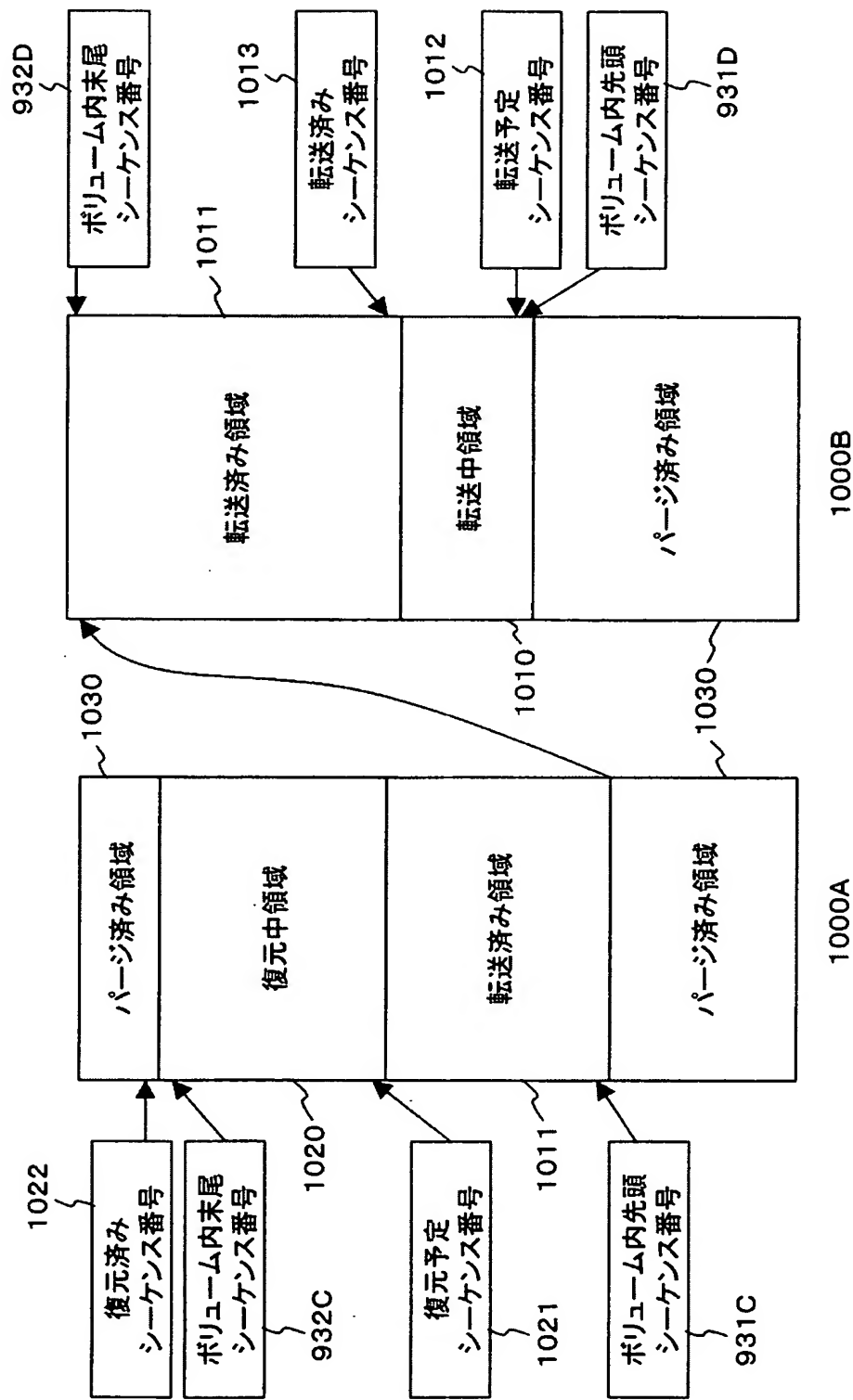
【図 7】

図 7



【図8】

図 8



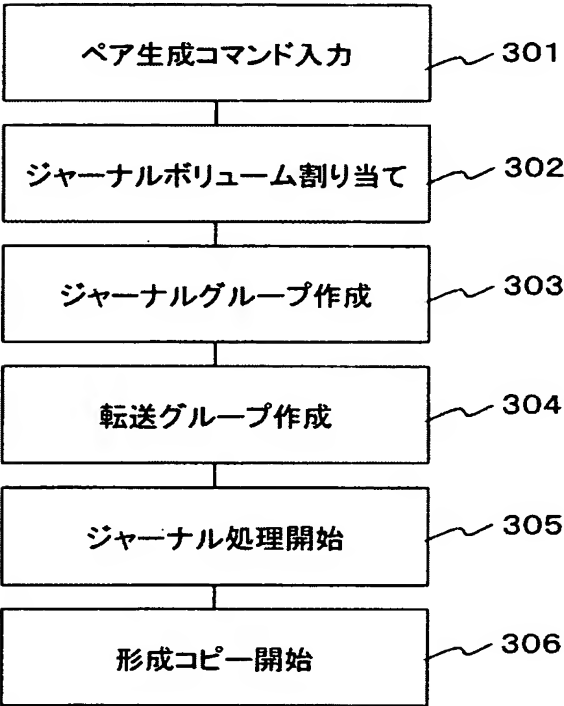
【図 9】

図 9

1110	切り替え判定有効フラグ			
1120	判定情報	ジャーナル量設定項目	有効フラグ	1121
			閾値条件	1122
		時間差設定項目	有効フラグ	1123
			閾値条件	1124
1130	条件外時動作			

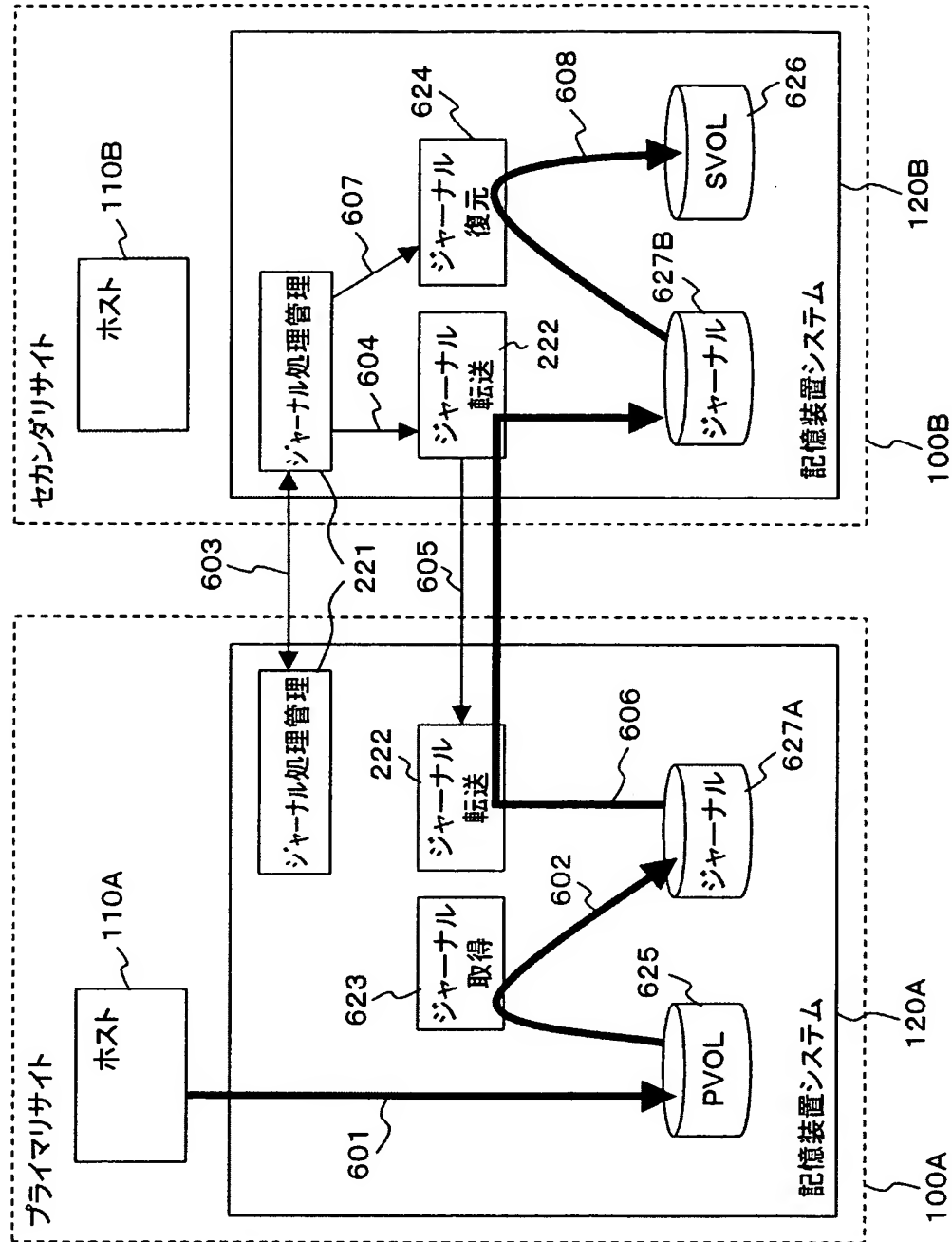
【図 1 0】

図 10



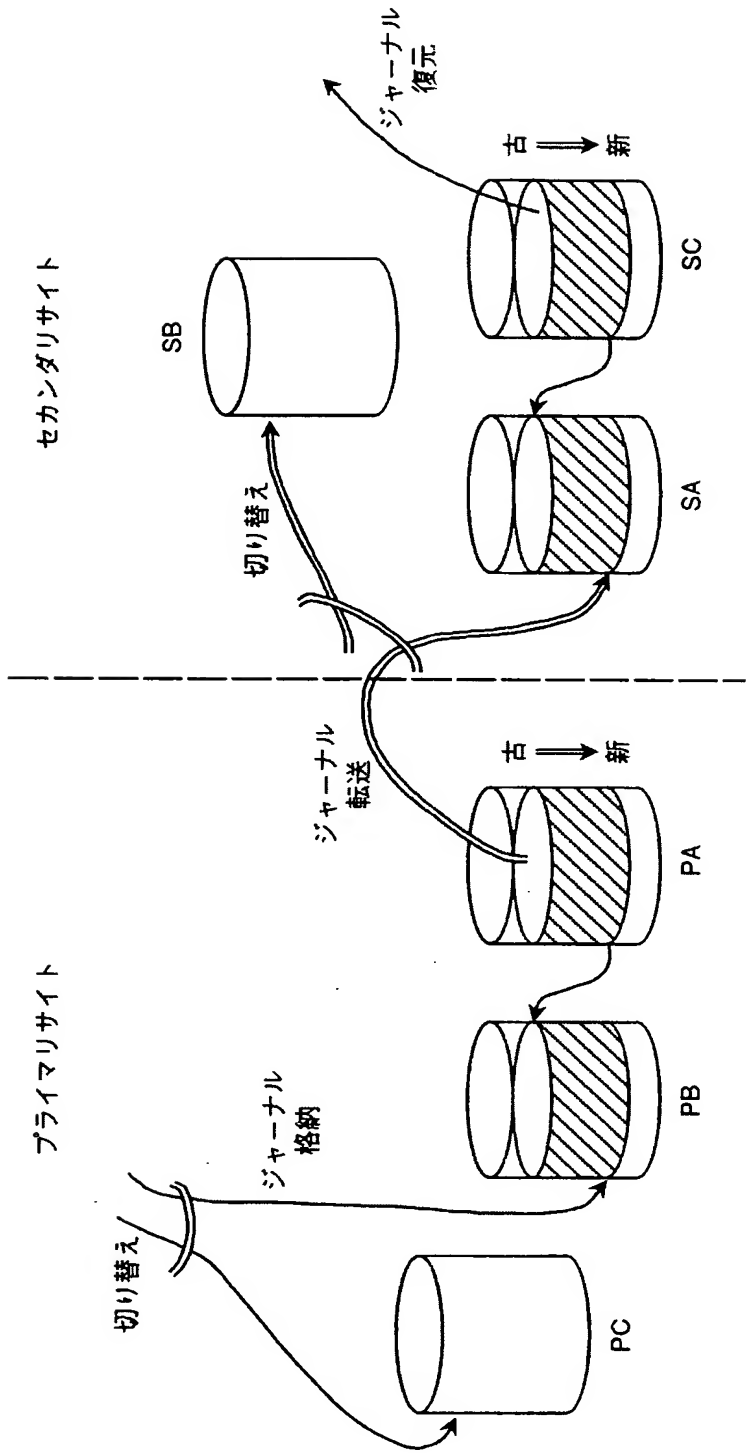
【図11】

図 11



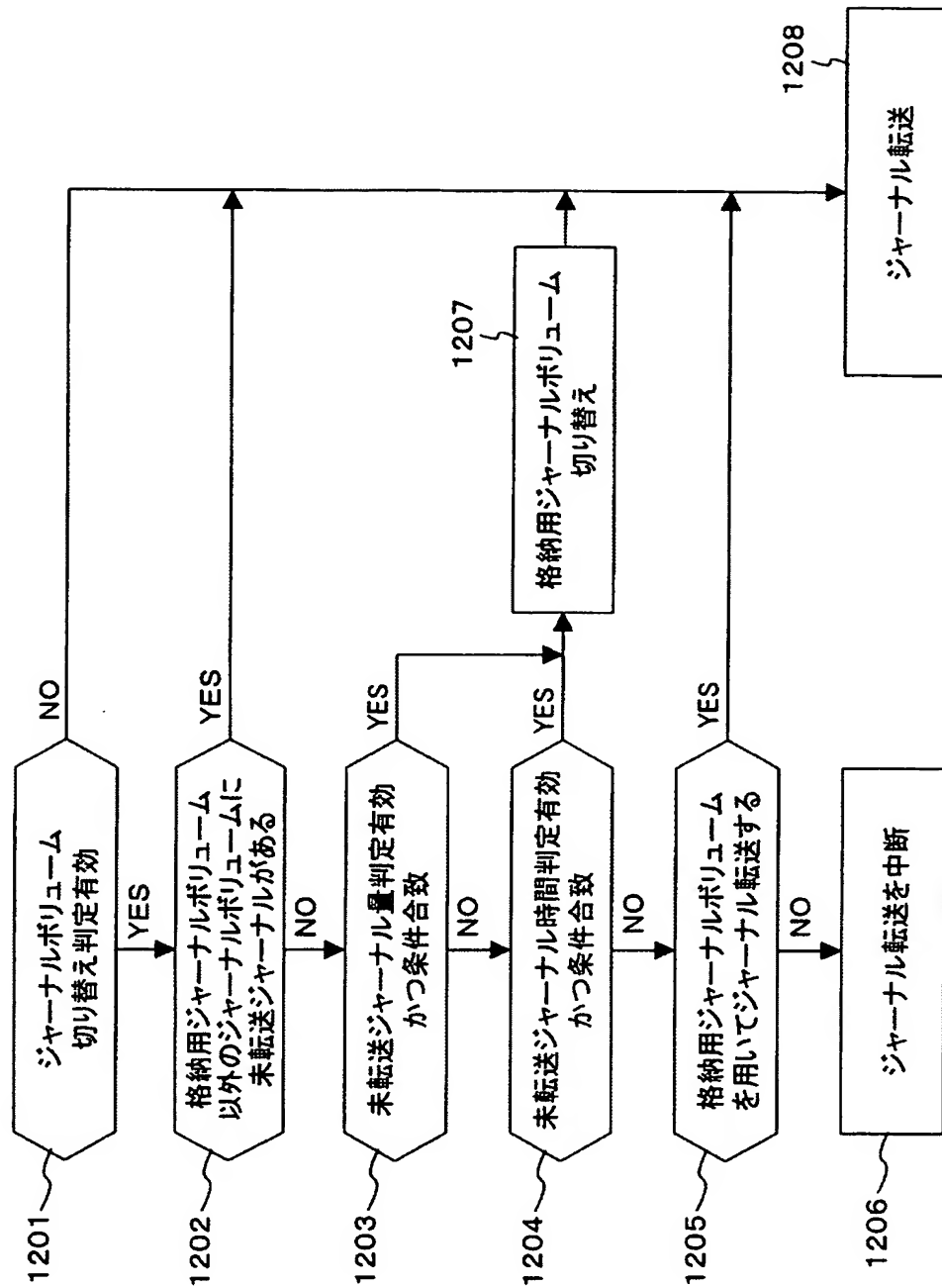
【図 12】

図 12



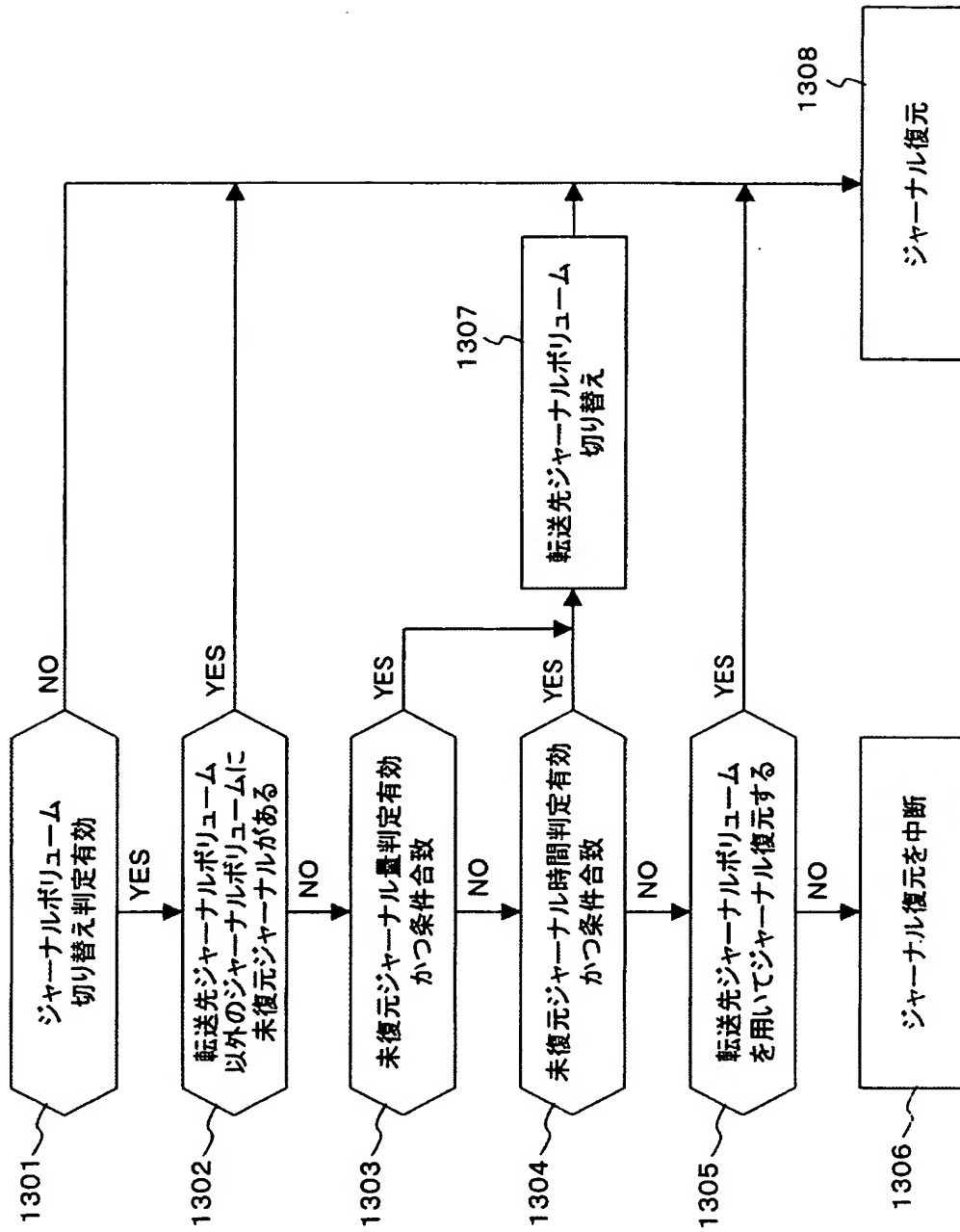
【図 13】

図 13



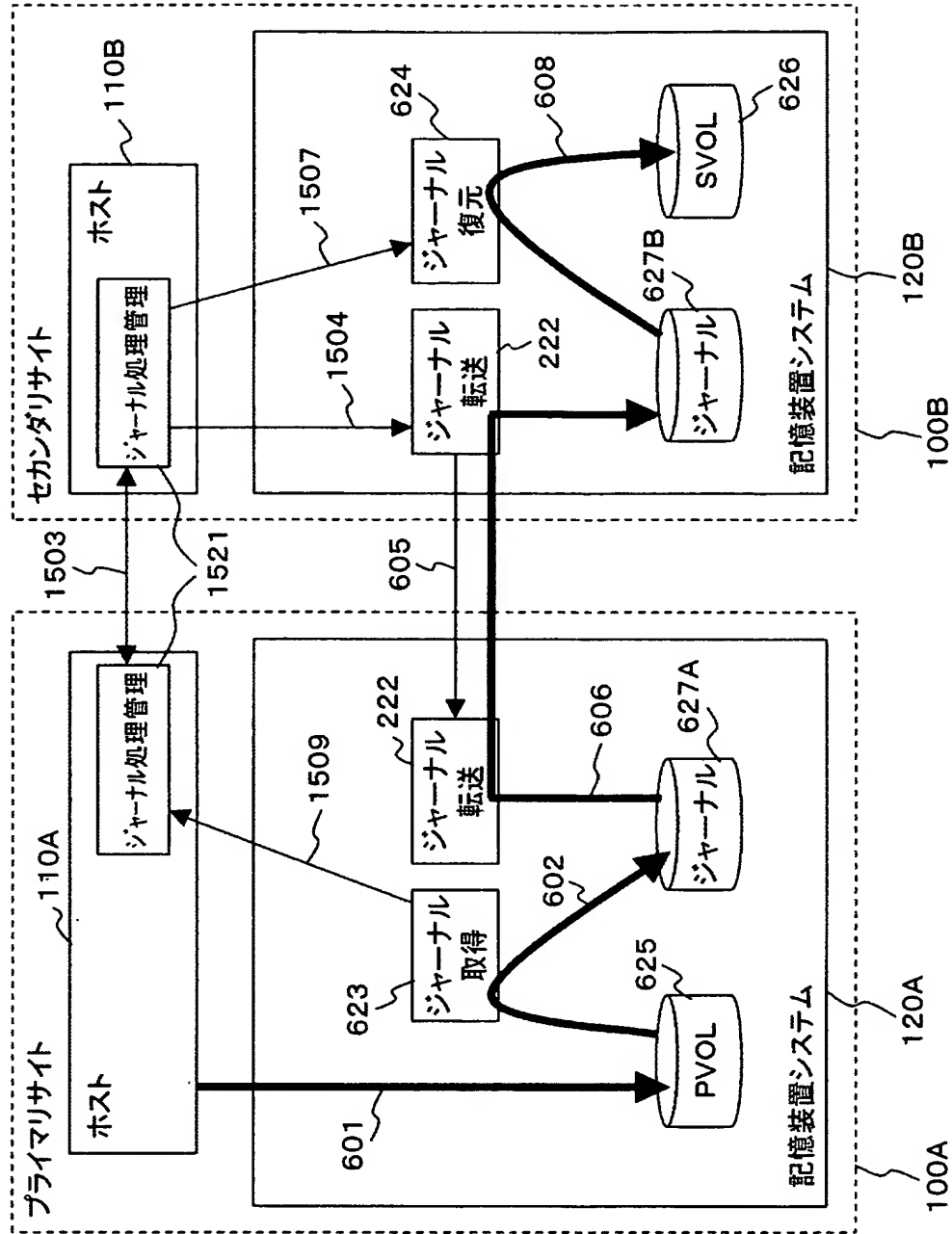
【図 14】

図 14



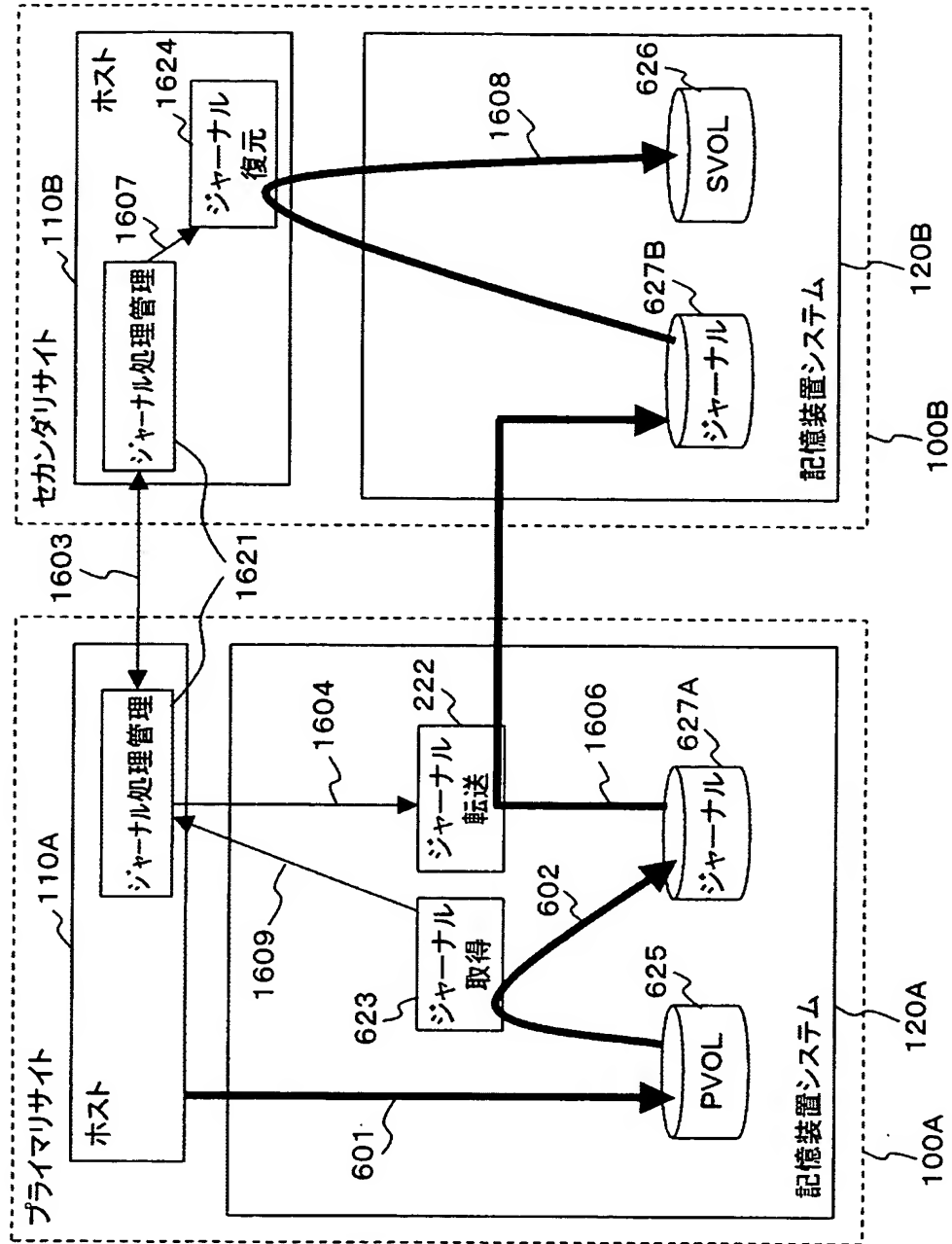
【図 16】

図 16



【図 17】

図 17



【書類名】 要約書**【要約】**

【課題】 データ処理システムにおいて、複数のサイトでデータの整合性を保証し、ホストとネットワークの負荷をかけることなく、しかも、データの更新復元に伴う特定の記憶装置への負荷集中を避けて、システムの処理能力を低下させないようにする。

【解決手段】 プライマリサイトと、セカンダリサイトとを有するデータ処理システムにおいて、記憶装置システム間を通信線により接続し、プライマリサイトにおいて、データの更新履歴をジャーナルとして記憶装置に記録し、ジャーナルをセカンダリサイトに前記通信線を介して転送する。転送の際に、プライマリサイトでは、ジャーナルを格納しているボリュームを切り替え、セカンダリサイトでは、ジャーナル転送先のボリュームを切り替え特定のボリュームに負荷が集中しないようにする。

【選択図】 図 1 1

特願 2 0 0 4 - 0 2 3 0 9 0

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 1 0 8]

1. 変更年月日	1 9 9 0 年 8 月 3 1 日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台 4 丁目 6 番地
氏 名	株式会社日立製作所